



Escola de Camins

Escola Tècnica Superior d'Enginyeria de Camins, Canals i Ports
UPC BARCELONATECH

**Comparativa de models univariants
per la predicció del tràfic de
contenidors. Aplicació als ports de
Barcelona, Tarragona i València.**

Treball realitzat per:

Basili Monturiol Giménez

Dirigit per:

Manel Grifoll Colls

Maribel Ortego

Grau en:

Enginyeria Civil

Barcelona, 22 de setembre de 2017

Departament d'Enginyeria Civil i Ambiental

TREBALL FINAL DE GRAU

RESUM

La predicció en el tràfic portuari és una eina molt útil per a les Autoritats Portuàries i els organismes de l'Estat per entendre com s'estan efectuant les evolucions en el tràfic i els moviments de contenidors en els ports sota el seu domini. El fet de disposar d'informació precisa respecte els futurs canvis permet anticipar les diferents accions a dur a terme i planificar una solució adequada per a cada tipus de port.

Aquest element previsor entra en valor dins del sistema portuari que es troba en constant evolució. Els nous paradigmes respecte als punts d'entrada i la seva connexió intermodal, així com els criteris de competitivitat, eficiència i productivitat, fan que els ports i les autoritats encarregades de la gestió hagin de buscar noves formes d'incrementar les possibilitats dels seus ports amb inversions molt costoses que requereixen de participacions amb entitats privades disposades a dur-les a terme en forma de concessions, ja sigui en terminals portuàries o en el desenvolupament d'un servei dins del sector.

Els mètodes utilitzats seran els univariants, característics per tenir una única variable, en aquest cas el temps. Es veurà quines són les propietats d'aquests mètodes de predicció basats en sèries temporals, dins dels quals l'estudi es centra en un model de regressió simple i l'anàlisi a través dels models ARIMA, i s'aplicaran als ports de Barcelona, Tarragona, València, i una anàlisi conjunta dels tres ports.

S'emfatitzarà en el fet que la tipologia de dades a tractar és molt important en el fet d'obtenir bones prediccions en els diferents models utilitzats, fet que es podrà observar en els resultats obtinguts tant dels dos períodes de verificació com en els gràfics d'estadístics que s'obtidran i en l'anàlisi dels errors dut a terme.

Es discutirà el fet que els models han funcionat millor quan les dades no han tingut grans canvis de tendència ni punts d'inflexió, ja que els ports on experimenten grans fluctuacions en el nombre de contenidors durant els períodes previs, durant, i posteriors a la crisi no han funcionat tan bé com en els altres.

Així mateix es farà present la forta tendència autoregressiva de les dades tractades, essent així els models amb ordres autoregressius els que finalment a nivell general s'han ajustat millor, amb unes característiques estadístiques prèviament acceptades.

Finalment, es conclourà que en tots els models acceptats com a vàlids la tendència en el nombre de TEU va a l'alça, reforçant la potencialitat d'aquesta zona superior del corredor mediterrani.

Paraules clau: Predicció, Sistema Portuari, Mètodes Univariants, Anàlisi de Sèries Temporals, Models ARIMA

ABSTRACT

Forecasting in traffic port is a useful tool for the Port Authorities and the government agencies for understanding how the traffic evolves in the ports under their domain. With this accurate data is possible to anticipate the different actions to be carried out and to plan an appropriate solution for each port.

This predetermining element comes into value within the port system that is constantly evolving. The new paradigms with respect to the gateways and their intermodal connection, as well as the criteria of competitiveness, efficiency and productivity, make the authorities to look for new ways to increase the possibilities of their ports with very expensive investments that require participation with private companies to carry them out, generally with concessions, either at port terminals or in providing some service into the port.

The methods used will be the univariate methods, characteristics for having a single variable, in this case the time. It will be seen what are the properties of these forecasting methods based on time-series, in which the study focuses on a simple regression model and the analysis through the ARIMA models, and will be applied to the ports of Barcelona, Tarragona, Valencia, and a joint analysis of the three ports.

It will be emphasized that the type of data to be treated is very important in the fact of obtaining good predictions in the different forecasting models used, which can be observed in the results obtained from both verification periods as well as in the statistics plots, and in the analysis of the errors carried out.

The fact that the models have worked better when the data has not had changes of tendency nor turning points, since ports where they experience large fluctuations in the number of containers during the periods prior to, during and after the economic recession have not worked as well as in others.

Moreover, the strong auto-regressive tendency of the data treated will be considered, being this models with auto-regressive orders the ones that have worked better, with a previously accepted statistical characteristics.

Finally, it will be concluded that in all models accepted as valid the tendency in the TEU number is rising, reinforcing the potential of this upper area of the Mediterranean corridor.

Keywords: *Forecasting, Port System, Univariate Methods, Time-series analysis, ARIMA Models*

ÍNDEX

ÍNDEX D'IL·LUSTRACIONS, TAULES I FIGURES.....	6
1. INTRODUCCIÓ I OBJECTIUS	11
2. SISTEMA PORTUARI.....	14
3. TÈCNiques DE PREDICCIÓ BASADES EN SÈRIES TEMPORALS.....	25
3.1. INTRODUCCIÓ ALS MODELS DE PREDICCIÓ	25
3.2. TÈCNiques GENERALS DE PREDICCIÓ	26
3.3. PRINCIPIS I RISCS DE LA PREDICCIÓ.....	28
3.4. MÈTODES I MODELS	32
3.4.1. Estadística bàsica i model de predicció més simple: la mitjana de la mostra	32
3.4.2. Model del camí aleatori (Random walk model).....	37
3.4.3. Prediccions amb mitjanes mòbils (Forecasting with moving averages).....	39
3.4.4. Models de regressió	45
3.4.5. Models ARIMA	50
3.5. PERÍODES DE VERIFICACIÓ	56
4. RESULTATS	58
4.1. DESCRIPCIÓ DE LES DADES	62
4.1.1. Barcelona.....	62
4.1.2. Tarragona	63
4.1.3. València	64
4.1.4. BTV.....	65
4.2. ESTIMACIONS I PREDICCIONS	65
4.2.1. Regressió simple	66
4.2.2. Model de la mitjana (ARIMA(0,0,0)+c)	68
4.2.3. Model del camí aleatori (ARIMA(0,1,0)+c).....	74
4.2.4. Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	81
4.2.5. Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c).....	87
4.2.6. Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c).....	93

4.2.7. Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	99
4.3. EXERCICIS DE VERIFICACIÓ 2010 I 2005	105
4.3.1. Regressió simple	105
4.3.2. Model de la mitjana (ARIMA(0,0,0)+c)	110
4.3.3. Model del camí aleatori (ARIMA(0,1,0)+c).....	114
4.3.4. Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	118
4.3.5. Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c).....	123
4.3.6. Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c).....	127
4.3.7. Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	132
4.3.8. Taules resum dels errors	136
5. DISCUSSIÓ	138
6. CONCLUSIONS.....	143
7. BIBLIOGRAFIA.....	146

ÍNDEX D'IL·LUSTRACIONS, TAULES I FIGURES

Il·lustració 1 Patró tipus mean model (Nau, 2014)	29
Il·lustració 2 Patró tipus Regressió lineal simple (Nau, 2014)	29
Il·lustració 3 Patró tipus Model del camí aleatori (Nau, 2014)	30
Il·lustració 4 Patró tipus Model estacional (Nau, 2014)	30
Il·lustració 5 Pautes d'ajust del model ARIMA en funció dels residuals i les autocorrelacions (Nau, 2014)	55

Taula 1 Taula d'errors obtinguda dels exercicis de verificació a les dades del port de Barcelona	136
Taula 2 Taula d'errors obtinguda dels exercicis de verificació a les dades del port de Tarragona	136
Taula 3 Taula d'errors obtinguda dels exercicis de verificació a les dades del port de València	137
Taula 4 Taula d'errors obtinguda dels exercicis de verificació a les dades conjuntes dels ports BTV	137
Taula 5 Taula tipificada per a l'anàlisi d'errors en les verificacions de les dades del port de Barcelona	138
Taula 6 Taula tipificada per a l'anàlisi d'errors en les verificacions de les dades del port de Tarragona	138
Taula 7 Taula tipificada per a l'anàlisi d'errors en les verificacions de les dades del port de València	139
Taula 8 Taula tipificada per a l'anàlisi d'errors en les verificacions de les dades conjuntes dels ports BTV	139

Figura 1 Evolució temporal del nombre de TEUS Barcelona	62
Figura 2 Evolució temporal del nombre de TEUS Tarragona	63
Figura 3 Evolució temporal del nombre de TEUS València	64
Figura 4 Evolució temporal del nombre de TEUS BTV	65
Figura 5 Regressió simple Barcelona	66
Figura 6 Regressió simple Tarragona	66
Figura 7 Regressió simple València	67
Figura 8 Regressió simple BTV	67
Figura 9 Model de la mitjana Barcelona	68
Figura 10 Model de la mitjana Barcelona gràfic de residuals	68
Figura 11 Model de la mitjana Barcelona gràfic d'autocorrelacions	69
Figura 12 Model de la mitjana Tarragona	69
Figura 13 Model de la mitjana Tarragona gràfic de residuals	70
Figura 14 Model de la mitjana Tarragona gràfic d'autocorrelacions	70

Figura 15 Model de la mitjana València	71
Figura 16 Model de la mitjana València gràfic de residuals	71
Figura 17 Model de la mitjana València gràfic d'autocorrelacions.....	72
Figura 18 Model de la mitjana BTV.....	72
Figura 19 Model de la mitjana BTV gràfic de residuals.....	73
Figura 20 Model de la mitjana BTV gràfic d'autocorrelacions	73
Figura 21 Model del camí aleatori Barcelona	74
Figura 22 Model del camí aleatori Barcelona gràfic de residuals	75
Figura 23 Model del camí aleatori Barcelona gràfic d'autocorrelacions	75
Figura 24 Model del camí aleatori Tarragona.....	76
Figura 25 Model del camí aleatori Tarragona gràfic de residuals.....	76
Figura 26 Model del camí aleatori Tarragona gràfic d'autocorrelacions	77
Figura 27 Model del camí aleatori València.....	77
Figura 28 Model del camí aleatori València gràfic de residuals.....	78
Figura 29 Model del camí aleatori València gràfic d'autocorrelacions	78
Figura 30 Model del camí aleatori BTV	79
Figura 31 Model del camí aleatori BTV gràfic de residuals	79
Figura 32 Model del camí aleatori BTV gràfic d'autocorrelacions.....	80
Figura 33 Model AR1 Barcelona	81
Figura 34 Model AR1 Barcelona gràfic de residuals	81
Figura 35 Model AR1 Barcelona gràfic d'autocorrelacions	82
Figura 36 Model AR1 Tarragona.....	82
Figura 37 Model AR1 Tarragona gràfic de residuals.....	83
Figura 38 Model AR1 Tarragona gràfic d'autocorrelacions	83
Figura 39 Model AR1 València.....	84
Figura 40 Model AR1 València gràfic de residuals.....	84
Figura 41 Model AR1 València gràfic d'autocorrelacions	85
Figura 42 Model AR1 BTV	85
Figura 43 Model AR1 BTV gràfic de residuals	86
Figura 44 Model AR1 BTV gràfic d'autocorrelacions.....	86
Figura 45 Model AR1 diferenciat Barcelona	87
Figura 46 Model AR1 diferenciat Barcelona gràfic de residuals	87
Figura 47 Model AR1 diferenciat Barcelona gràfic d'autocorrelacions	88
Figura 48 Model AR1 diferenciat Tarragona.....	88
Figura 49 Model AR1 diferenciat Tarragona gràfic de residuals.....	89
Figura 50 Model AR1 diferenciat Tarragona gràfic d'autocorrelacions	89
Figura 51 Model AR1 diferenciat València.....	90
Figura 52 Model AR1 diferenciat València gràfic de residuals.....	90
Figura 53 Model AR1 diferenciat València gràfic d'autocorrelacions	91
Figura 54 Model AR1 diferenciat BTV	91
Figura 55 Model AR1 diferenciat BTV gràfic de residuals	92
Figura 56 Model AR1 diferenciat BTV gràfic d'autocorrelacions.....	92
Figura 57 Model AR2 diferenciat Barcelona	93

Figura 58 Model AR2 diferenciat Barcelona gràfic de residuals	93
Figura 59 Model AR2 diferenciat Barcelona gràfic d'autocorrelacions	94
Figura 60 Model AR2 diferenciat Tarragona.....	94
Figura 61 Model AR2 diferenciat Tarragona gràfic de residuals.....	95
Figura 62 Model AR2 diferenciat Tarragona gràfic d'autocorrelacions	95
Figura 63 Model AR2 diferenciat València.....	96
Figura 64 Model AR2 diferenciat València gràfic de residuals.....	96
Figura 65 Model AR2 diferenciat València gràfic d'autocorrelacions	97
Figura 66 Model AR2 diferenciat BTV	97
Figura 67 Model AR2 diferenciat BTV gràfic de residuals	98
Figura 68 Model AR2 diferenciat BTV gràfic d'autocorrelacions.....	98
Figura 69 Model d'ajustament exponencial Barcelona.....	99
Figura 70 Model d'ajustament exponencial Barcelona gràfic de residuals.....	99
Figura 71 Model d'ajustament exponencial Barcelona gràfic d'autocorrelacions	100
Figura 72 Model d'ajustament exponencial Tarragona	100
Figura 73 Model d'ajustament exponencial Tarragona gràfic de residuals ...	101
Figura 74 Model d'ajustament exponencial Tarragona gràfic d'autocorrelacions	101
Figura 75 Model d'ajustament exponencial València	102
Figura 76 Model d'ajustament exponencial València gràfic de residuals	102
Figura 77 Model d'ajustament exponencial València gràfic d'autocorrelacions	103
Figura 78 Model d'ajustament exponencial BTV	103
Figura 79 Model d'ajustament exponencial BTV gràfic de residuals	104
Figura 80 Model d'ajustament exponencial BTV gràfic d'autocorrelacions ...	104
Figura 81 Regressió simple Barcelona verificació 2010.....	105
Figura 82 Regressió simple Barcelona verificació 2005.....	106
Figura 83 Regressió simple Tarragona verificació 2010	106
Figura 84 Regressió simple Tarragona verificació 2005	107
Figura 85 Regressió simple València verificació 2010	107
Figura 86 Regressió simple València verificació 2005	108
Figura 87 Regressió simple BTV verificació 2010	108
Figura 88 Regressió simple BTV verificació 2005	109
Figura 89 Model de la mitjana Barcelona verificació 2010	110
Figura 90 Model de la mitjana Barcelona verificació 2005	110
Figura 91 Model de la mitjana Tarragona verificació 2010	111
Figura 92 Model de la mitjana Tarragona verificació 2005	111
Figura 93 Model de la mitjana València verificació 2010	112
Figura 94 Model de la mitjana València verificació 2005	112
Figura 95 Model de la mitjana BTV verificació 2010.....	113
Figura 96 Model de la mitjana BTV verificació 2005.....	113
Figura 97 Model del camí aleatori Barcelona verificació 2010	114

Figura 98 Model del camí aleatori Barcelona verificació 2005	114
Figura 99 Model del camí aleatori Tarragona verificació 2010.....	115
Figura 100 Model del camí aleatori Tarragona verificació 2005	115
Figura 101 Model del camí aleatori València verificació 2010	116
Figura 102 Model del camí aleatori València verificació 2005	116
Figura 103 Model del camí aleatori BTV verificació 2010	117
Figura 104 Model del camí aleatori BTV verificació 2005	117
Figura 105 Model AR1 Barcelona verificació 2010.....	118
Figura 106 Model AR1 Barcelona verificació 2005.....	119
Figura 107 Model AR1 Tarragona verificació 2010	119
Figura 108 Model AR1 Tarragona verificació 2005	120
Figura 109 Model AR1 València verificació 2010	120
Figura 110 Model AR1 València verificació 2005	121
Figura 111 Model AR1 BTV verificació 2010	121
Figura 112 Model AR1 BTV verificació 2005	122
Figura 113 Model AR1 diferenciat Barcelona verificació 2010	123
Figura 114 Model AR1 diferenciat Barcelona verificació 2005	123
Figura 115 Model AR1 diferenciat Tarragona verificació 2010	124
Figura 116 Model AR1 diferenciat Tarragona verificació 2005	124
Figura 117 Model AR1 diferenciat València verificació 2010	125
Figura 118 Model AR1 diferenciat València verificació 2005	125
Figura 119 Model AR1 diferenciat BTV verificació 2010.....	126
Figura 120 Model AR1 diferenciat BTV verificació 2005.....	126
Figura 121 Model AR2 diferenciat Barcelona verificació 2010	127
Figura 122 Model AR2 diferenciat Barcelona verificació 2005	128
Figura 123 Model AR2 diferenciat Tarragona verificació 2010	128
Figura 124 Model AR2 diferenciat Tarragona verificació 2005	129
Figura 125 Model AR2 diferenciat València verificació 2010	129
Figura 126 Model AR2 diferenciat València verificació 2005	130
Figura 127 Model AR2 diferenciat BTV verificació 2010.....	130
Figura 128 Model AR2 diferenciat BTV verificació 2005.....	131
Figura 129 Model d'ajustament exponencial simple Barcelona verificació 2010	132
Figura 130 Model d'ajustament exponencial simple Barcelona verificació 2005	132
Figura 131 Model d'ajustament exponencial simple Tarragona verificació 2010	133
Figura 132 Model d'ajustament exponencial simple Tarragona verificació 2005	133
Figura 133 Model d'ajustament exponencial simple València verificació 2010	134
Figura 134 Model d'ajustament exponencial simple València verificació 2005	134

Figura 135 Model d'ajustament exponencial simple BTV verificació 2010 135

Figura 136 Model d'ajustament exponencial simple BTV verificació 2005 135

1. INTRODUCCIÓ I OBJECTIUS

La predicció del tràfic portuari, igual que en tots els altres sectors on pot ser aplicada la predicció, ofereix una eina amb la qual anticipar-se a canvis que es produiran en un futur proper.

El fet de tenir informació fiable i obtinguda sota mètodes estadísticament correctes en un context real ens permetrà dur a terme una planificació global més acurada, i serà possible tenir en compte els factors que s'hauran determinat com a clau per al futur del sector estudiat.

Per entendre els objectius que es plantejaran en aquest treball en relació al sistema portuari es clau entendre l'instrument en el qual ens basarem per a les prediccions. Aquest serà el TEU (*Twenty-Foot Equivalent Unit*) que es correspon amb la unitat de mesura de capacitat inexacta del transport marítim expressada en contenidors.

L'evolució dels TEU dins del sistema portuari es coneix amb el nom de containerització i la seva evolució es pot descriure com una onada del tràfic de TEU en relació amb el temps, que es caracteritza per un període d'adopció de la unitat de mesura, un canvi de pendent positiu que s'anomena acceleració al qual segueix un període de creixement màxim on tindrem la pendent més gran, i finalment seguirà una fase de maduresa en que la pendent torna a ser suau (Guerrero i Rodrigue, 2014).

La predicció d'aquests TEU en relació amb la seva evolució amb el temps es pot dur a terme a través de diferents mètodes, cadascun amb les seves particularitats.

Disposar de prediccions fiables en aquest aspecte és de gran ajuda per a l'evolució dels ports que estan destinats bàsicament al moviment de contenidors. Les prediccions són doncs de gran interès per a les Autoritats Portuàries a l'hora de prendre decisions que afectin a la competitivitat i l'eficiència dels seus serveis.

En l'entorn portuari actual els ports assoleixen un nou estatus en la xarxa global de transport dins dels quals estan immersos, fins al punt que cada cop més són considerats una part de la cadena de subministrament de les empreses multinacionals (Robinson, 2002). D'aquesta manera la competitivitat i l'eficiència que puguin tenir poden ser factors clau en la seva elecció.

Entendre les fluctuacions en el tràfic portuari pot portar a fer inversions més adequades per la importància que tingui el port concret dins de tal cadena de subministrament.

En aquest nou entorn canviant cada cop són més comuns les participacions entre les entitats públiques i empreses privades per tal de fer front a tals inversions, generalment en forma de concessions (Saurí, 2007). Aquests acords van destinats a ampliar les capacitats dels ports per atraure negoci. En aquests

acords el paper de la predicció és clau per tal de que la proposta d'inversió en un o altre port per part del capital privat resulti atractiva a l'hora del retorn que aquestes puguin tenir.

En el sistema actual, els ports no sol aporten valor a la cadena sinó també a la indústria del territori interior que es capaç d'abastir. D'aquesta manera el fet que les prediccions per als ports d'un determinat territori siguin propícies afecten indirectament al moviment econòmic de les regions que tenen interconnexió amb el port, ja que serien els beneficiats de poder disposar de noves infraestructures amb el capital captat de les inversions, com ara una major facilitat en el transport interior degut a millores en la xarxa intermodal (Van Link, 1998).

Amb tot, s'estableix un joc d'equilibri entre les Autoritats Portuàries i l'Estat que tenen necessitat de captar capital del qual no disposen per ampliar les diferents infraestructures de les que en són propietaris, o per fer front a certs serveis dels ports, i les entitats privades que busquen beneficis de les aportacions que puguin fer i de l'entrada al marc logístic dels ports. Una eina de gran importància per a que siguin duts a terme els acords serà presentar un pla en relació a les previsions de creixement que tingui la zona en concret, elaborats a partir de prediccions.

Entrant en matèria de les prediccions, aquestes poden ser bàsicament a partir de mètodes qualitatius i quantitatius. De forma simplificada es pot entendre per qualitatius les opinions d'experts en relació amb el tema que estem tractant, i per quantitatius tots aquells mètodes que utilitzen les diverses eines de càlcul per arribar a les prediccions (*Port Traffic Forecasting Tool*, 2015). Bàsicament en distingim les que utilitzen indicadors econòmics, poblacionals, dades concretes del sector tractat... i les que es basen en l'anàlisi de sèries temporals.

En aquesta memòria les prediccions es basaran en l'anàlisi de sèries temporals per tal d'analitzar l'evolució dels TEU en relació a les dades històriques de les quals disposem en els casos tractats, és per això que els mètodes utilitzats s'anomenen univariants, ja que la variable a la qual es fa referència per estudiar els diferents models i mètodes serà el temps.

Tals mètodes ja han estat utilitzats en estudis que fan referència a ports asiàtics (Peng, 2009) i els resultats obtinguts fan pensar que són una bona eina a tenir en compte a l'hora d'elaborar les prediccions, de manera que pot ser interessant la seva aplicació a ports del nostre territori.

Els ports que s'analitzaran amb aquests mètodes seran el de Barcelona, el de Tarragona i el de València, als que s'hi referirà com BTV.

Tot i que inicialment l'estudi estava pensat per l'anàlisi singular del cas de Barcelona, s'ha decidit ampliar aquest estudi als ports de Tarragona i València pel fet que també s'inclouen dins de les 10 Autoritats Portuàries amb més import de xifra de negocis ICN de l'Estat, i és interessant l'anàlisi de cara a posar en valor tot el que esdevé aquesta zona superior del corredor mediterrani a nivell

de transport cap a l'interior i l'exterior peninsular. Aquests ports s'analitzaran tant de forma individual com finalment en conjunt tenint en compte tots els TEU que s'hi mouen entre els tres.

Els objectius d'aquest treball són realitzar l'estudi teòric dels diferents models d'anàlisi de sèries temporals que es presentaran en el quart apartat de la memòria, per posteriorment dur a terme una comparativa pràctica de la fiabilitat d'aquests models i extreure'n les màximes conclusions possibles al respecte.

Es partirà com a base de la comparativa entre un model simple i de poca complexitat, i s'anirà revisant en comparació dels altres models de complexitat superior.

Amb tot, es podrà veure si aquests tipus de mètodes de predicció són vàlids o acceptables pels ports en que els analitzem, valorant la seva bondat no tan sols en l'ajust de les prediccions que en farem sinó també basant-nos en el rigor estadístic que cada model haurà de tenir per tal de ser considerat acceptable i vàlid.

Independentment dels resultats obtinguts, també es farà una valoració global de quan poden funcionar aquests tipus de models, quan no, quines dades o quin tipus de dades pot fer que els models fallin, com podem ajustar-los, etc. Així mateix, també es valorarà les prediccions que resultin dels mètodes que hagin estat considerats amb millor ajust, per tal de veure si s'espera decreixement, estancament o creixement dels TEU en la zona analitzada.

L'estructura en el desenvolupament d'aquest estudi es compondrà d'una breu explicació del sistema portuari actual, tant a nivell global, europeu, com espanyol, que ampliarà molts dels conceptes introduïts en aquest primer apartat. S'hi desenvoluparan conceptes més de caire general pel que fa als ports i a l'entorn d'aquests i s'aprofundirà més en el concepte actual de port, dels objectius d'aquests, de la forta competència en un entorn cada cop més globalitzat, de les Autoritats Portuàries, etc.

Posteriorment, s'explicaran els conceptes de predicció en si i dels diferents tipus de models i tècniques en que es pot classificar l'anàlisi de sèries temporals. Amb tota la teoria desenvolupada, es passarà a l'estudi pràctic on es presentaran el seguit de resultats que hem anat obtenint, separant-los en els diferents períodes de verificació tractats i les prediccions dels diferents models.

Finalment una vegada processats els diferents resultats obtinguts es passarà a la discussió de les solucions i a l'elaboració de les diferents conclusions que se'n podran extreure.

2. SISTEMA PORTUARI

La no existència d'unes normes acceptades per unanimitat per guiar les accions de les companyies multinacionals fa que s'emfatitzi tant en els moviments estratègics com en la confluència logística, la intermodalitat i la integració de les xarxes de transport.

Els ports estan vinculats als *hinterlands* o punts de connexió amb l'interior pròxims, i la selecció de les rutes i dels ports depèn de la cadena global de subministrament més que de les pròpies característiques dels recintes portuaris. En aquest context, la predicció del tràfic portuari ens serveix com una eina molt útil per entendre aquesta nova situació dels ports i anticipar-nos a pròxims canvis.

És necessari doncs entendre el sistema portuari i la seva evolució per a poder plantejar-nos diverses situacions i fer-ne a posteriori les apreciacions necessàries.

Es pot afirmar que les pròpies funcions portuàries posseeixen i tenen un paper essencial en les economies territorials.

Rimmer (2007), en la seva reflexió sobre les evolucions portuàries diferenciava els següents processos: primerament existeixen els ports dispersos o atomitzats al llarg de la costa, amb un caràcter molt autònom; posteriorment, es visualitzen línies de connexió portuària entre ports pròxims; més tard, són molt notòries les dinàmiques de interconnexió i de concentració portuària; després, es consolida una tendència cap a la centralització portuària; i, finalment, es constaten dinàmiques de desconcentració i descentralització portuària degut a motius com la congestió dels ports més grans, entre d'altres.

Aquests canvis venen acompanyats d'uns vaixells cada cop més grans, amb una complexitat de cadenes de subministrament cada cop superior, que juntament amb la propagació i desenvolupament dels avenços tecnològics van ser factors clau per a la formació de *hubs off-shore* o centres logístics propers al litoral, tal i com defensen Notteboom i Rodrigue (2005).

Aquests *hubs* es poden considerar com la sisena fase de la periodificació de les dinàmiques portuàries, també anomenada per alguns autors "regionalització".

Aquesta discussió va obrir el debat sobre la competitivitat portuària, Notteboom i Rodrigue (2005), seguint la línia de raonaments de Robinson (2002), interpreten que els ports són simples elements de les cadenes de subministrament de les corporacions multinacionals. De fet, Robinson ho defineix de la manera següent: un port "és un proveïdor de serveis a tercers que intervé en les cadenes de subministrament com a una entitat enfocada al mercat; o també, que el port competeix amb altres ports com una companyia centrada en el mercat o en la entitat que forma part de la constel·lació de cadenes de subministrament enfocada al port".

D'acord amb aquesta definició es pot suposar el fet que els que competeixen són les cadenes, no els ports individuals. Sota aquesta perspectiva, els esquemes portuaris, en sentit global, amplien les economies d'escala a les terminals interiors, i, conseqüentment, les seves respectives zones logístiques.

En la nova visió en relació a les terminals, la competitivitat es planteja doncs sobre la base de millorar i generar valor afegit a les cadenes de subministrament.

A grans trets, els elements que més influeixen en la competència portuària podríem dir que són els següents:

- Estructures de govern, propietat i privatització.
- Actuació inter-portuària en termes de competitivitat i en relació a la productivitat i l'eficiència.
- Opcions de competència i cooperació entre ports.
- Classificació de les polítiques portuàries.
- Indicadors a través de les decisions relatives a l'elecció portuària.

Tot i que posteriorment es desenvoluparà més algun dels punts, alguns autors han estudiat algunes de les relacions mencionades en termes d'eficiència i potenciament del sistema portuari per a fer les seves mesures de la competitivitat.

Per exemple, Tongzon i Heng Wu (2005), van investigar la relació quantitativa entre la estructura de la propietat i la eficiència portuària, centrant-se en el moviment de contenidors per hora de treball del vaixell. Per la seva part, Talley (2006, 2007), havia analitzat les comparacions entre les produccions totals i les produccions òptimes, buscant uns models d'ajust i de projecció futura. Finalment, Thai i Grevall (2005), identifiquen els models d'anàlisi centrant-se en la divisió de les responsabilitats entre els diferents agents governamentals.

La cooperació i la competitivitat portuàries fan referència a com es preveu la resposta de les autoritats portuàries en front dels canvis del mercat. Cada vegada són majors les aliances portuàries enteses en termes tant de fusió com de cooperació empresarial. Song (2003), insitua en que les aliances cooperatives entre els ports de Hong Kong i Shenzhen eren molt útils. Altres autors com Van Klink i Van den Berg (1998), criden l'atenció sobre el fet que els *gateways* o portes d'entrada estan excel·lentment posicionades per potenciar el transport intermodal a Europa; per lo que tant el potenciament del *hinterland* portuari com la introducció (des de les aportacions de les Autoritats Portuàries) de trens de enllaços entre grans àrees econòmiques fan que els ports mantinguin i augmentin les seves quotes de mercat i millorin l'eficiència del sistema de transport. Haralambides (2002), tracta la qüestió des de la perspectiva de les tarifes; Pallis i Vaggelas (2005), analitzen a través de les accions de la Comissió Europea les diferents opcions competitives en el que fa referència al

funcionament dels serveis portuaris; i, McCalla (1999), fa menció als ajustos locals en referència als operadors que estan presents o que és possible atreure a les terminals i poden fer ús de la intermodalitat.

Al referenciar els factors de selecció portuària, els elements bàsics en principi eren quatre: distància oceànica, distància terrestre, freqüència de sortides dels vaixells, i capacitat d'aquests. També es poden afegir posteriorment els serveis logístics i les estratègies de companyies marítimes i operadors de terminals. Recentment, també s'ha de considerar el poder de quota dels *hinterlands* solapats.

Analitzant els serveis portuaris, la major globalització econòmica i la consideració dels ports com elements de la cadena de subministrament impulsen noves reflexions per part dels estudiosos de l'economia portuària.

El que ha significat un canvi en les dinàmiques portuàries es la nova situació i les noves relacions dels proveïdors dels serveis portuaris, que es converteix en l'enfocament central d'una gran part de les preocupacions de la gestió portuària.

La globalització ha suposat, en primer terme, una nova participació del públic i el privat en el sector portuari; en segon lloc, l'emergència d'uns serveis més sofisticats i integrats per part dels operadors; i finalment, unes noves relacions entre proveïdors i usuaris dels mencionats serveis portuaris.

En aquest nou context, la rellevància principal es la presència d'un major nombre d'actors. La containerització (Guerrero i Rodrigue, 2014) ha facilitat l'expansió logística i ha reclamat la presència de vaixells de major grandària i tecnologia, que implica una major demanda d'inversions i majors equipaments en molls i terminals, i també d'infraestructures per una gestió eficient de les operacions que redueixin els temps d'estada del vaixell en el port, de les operacions d'atrada, així com els augments en la productivitat de l'atrada en si.

Per aconseguir-ho els ports proporcionen nous serveis en la línia de la innovació, coneixement i planificació descentralitzades. A més, els conceptes d'entrada (*gateway*) són substituïts pel de centre logístic (*hub*), per proporcionar nous serveis complementaris i de coordinació de tràfics. Un altre punt seria la descentralització de la provisió dels serveis portuaris lligada a la connexió dels serveis portuaris amb àrees semi-autònomes o complementàries i afegides per una major captació i atracció de fluxos de mercaderies.

Aquest desenvolupament es basa principalment en l'especialització i la flexibilització (Saurí, 2007). L'especialització ens porta a treballar en forma de xarxa, centrant els objectius en la propietat i en el subministrament d'actius. Per la seva part, a través de la flexibilització, actuem de cara al mercat i a l'usuari, desenvolupant accions que generen un major valor afegit, com els centres logístics, el que estimula la presència de múltiples negocis diferenciats.

La actual estructura de govern portuari aporta una gran pluralitat de combinacions, mostrant la coexistència pública/privada. Tradicionalment, en la majoria dels països les Autoritats Portuàries eren les propietàries i les explotadores de les instal·lacions portuàries, no obstant, no estaria tan clar que aquesta sigui la manera més eficient d'organitzar aquesta indústria. El major protagonisme de la iniciativa privada és degut entre d'altres factors a la necessitat cada cop major de fons de finançament d'infraestructures i equipaments ja esmentat, a la necessitat de reduir dèficits públics, també a la idea de que la iniciativa privada esta més capacitada per obtenir majors productivitats per tenir aquesta una major capacitat d'adaptació a les necessitats, a la contribució al creixement del comerç, i a l'adquisició d'experiència en la gestió portuària.

Per tant, s'està tendint a una coparticipació pública i privada que permet als ports adaptar-se a l'entorn descrit de cada cop major competència i a una menor disponibilitat de fons de finançament públics.

Des d'aquesta perspectiva, principalment es poden distingir dos modalitats extremes d'organització portuària en el món:

- Model *landport*: el privat es fa càrrec del màxim possible de les activitats inscrites en l'àmbit portuari, essent l'Autoritat Portuària l'organisme propietari de tot l'immobilitzat que alquila o concessionaria el privat. En són exemples la gran majoria dels ports d'Europa, Estats Units i Canadà.
- Model de gestió integral: l'Autoritat Portuària s'encarrega directament de gestionar la totalitat dels serveis portuaris. En són exemples alguns ports africans.

Si optem per la participació del capital privat, podem segmentar-ho en dos grans categories principals:

- Privatització parcial: seria el model portuari *landlord* en el que els serveis relatius a les operacions portuàries serien privatitzades, però el port conservaria la propietat sobre la infraestructura.
- Privatització absoluta: es produiria un traspàs total a la iniciativa privada de les competències que fins ara eren del servei públic, incloent-hi la propietat de la infraestructura, planejament, inversió i gestió.

Per la combinació pública-privada hi ha diverses alternatives, la elecció de les quals dependrà del volum de l'activitat, de les circumstàncies de partida i del tipus de servei considerat. Les més significatives són: *Build, Operate and Own* (BOO), *Built/Rehabilitates, Operate and Transfer* (BOT o ROT, es tracta del cas típic de les concessions), *Joint-ventures*, *Leasing*, *Licensing* (autoritzacions) i Contractes de gestió entre d'altres.

Dins d'aquest grup, les concessions són el tipus particular de fórmula de participació privada en la gestió de ports més utilitzada en la majoria dels ports europeus, principalment en les terminals de contenidors.

Sota aquest model de gestió de ports, tipus *landport*, gran part dels actius es mantenen sota la propietat de l'Autoritat Portuària, essencialment l'infraestructura, transferint a la iniciativa privada tot el risc operatiu i financer de la gestió portuària.

Més concretament, la concessió es un contracte entre un operador privat i l'Autoritat Portuària en virtut del qual aquest segon transfereix al primer els drets de l'explotació portuària en qüestió, encara que manté la propietat de la major part dels actius, essencialment de la infraestructura i els terrenys. Usualment el concessionari ha de fer front a la rehabilitació i a la construcció.

Cal tenir en compte que existeix un *trade-off* (compensació/equilibri entre risc i beneficis), entre inversió i disponibilitat de informació per regular, entenent per regulació a la intervenció pública en el funcionament dels mercats en el sentit de controlar o marcar tarifes, beneficis o ingressos; controlar l'entrada o sortida del mercat; i controlar que les pràctiques del sector s'ajustin a la lliure competència, amb els objectius d'assegurar la competència entre els diversos operadors del port, controlar els monopolis tant existents com els potencials, i evitar pràctiques anticompetitives. Un altre factor a tenir en compte en la definició del termini de la concessió és la necessitat de recuperació de les inversions realitzades pel privat.

Amb tot, hi ha autors que sostenen que no sol es pot augmentar l'eficiència amb la privatització total, sinó que hi hauria altres vies que podrien ser: la modernització de l'administració i gestió portuària per part del sector públic, la liberalització dels serveis portuaris, la comercialització i la corporativització (Autoritat Portuària amb status d'empresa privada).

Per tant, el que queda clar és que la reconfiguració dels actors que conformen el sistema portuari ha estat molt evident en els últims anys. D'una part, les companyies marítimes utilitzen més vaixells amb majors capacitats, conformant un "oligopoli atípic" (Allix, 2006). Per una altra part, les terminals portuàries avancen en la quota de mercat mundial, internacionalitzant-se i configurant una nova xarxa d'espais de carga, descarrega i transbord. La major part de les grans companyies estan desenvolupant diferents formes de cooperació, reforçant les aliances i els acords amb l'objectiu de cobrir les rutes principals.

Es pot plantejar quin és o quin ha de ser el nou entorn portuari. En aquest sentit els ports estan cada cop més considerats com clústers, això és, estan formats per corporacions heterogènies amb negocis centrals diferenciats. A mesura que canvia l'entorn dels ports, els actors redefeixen les seves estratègies i l'abast de les seves activitats. El significat del primer exposat és el següent:

- Major participació dels actors en el mercat dels serveis portuaris. Això suposa l'existència de més d'un proveïdor de serveis en el mateix port, amb l'objectiu de perfeccionar les funcions d'especialització, innovació i flexibilitat.

- La participació del sector públic en la indústria portuària va disminuint en favor del sector privat que entra en alguna de les formes participatives comentades.
- Reducció de les barreres d'entrada a l'activitat, el que implica que els agents econòmics adoptin valors i pràctiques diferencials.

Una segona línia d'argumentacions es deriva de la pròpia coexistència dels actors públics i privats. D'una part es consideren alguns serveis portuaris com béns públics gestió de la qual ha de ser assumida pel sector públic. D'una altra part, es considera que certs serveis públics no haurien de ser responsabilitat exclusiva de les autoritats públiques. Això suposaria l'existència d'alguna de les combinacions de govern entre el sector públic i privat ja comentades, però, tenint en compte que és difícil estimar l'interval òptim dels sectors públics i privat en la seva participació.

El mercat de transport marítim de línies regulars es caracteritza per ser un oligopoli amb una gran tendència a la concentració empresarial y per assolir economies d'escala tant en el que fa referència als canvis organitzatius com als canvis tecnològics. Les recents onades de fusions i adquisicions d'empreses marítimes han estat molt intenses en els últims anys, fet que fa que les empreses més grans aglutinin la major part del mercat. De la mateixa manera, es fa present com les entitats que ofereixen serveis d'estiba i desestiba en els ports es poden subdividir en tres categories: els transportistes oceànics, els operadors internacionals de transports, i els conglomerats privats. El desenvolupament de nous models de transport estimulen un entorn logístic innovador, basat en la creació de valor i buscant una integració corporativa. Són, doncs, els grans transitaris els que ofereixen solucions integrades en tota la cadena de subministrament. Aquestes empreses competeixen dins del port de forma directa amb els proveïdors de serveis portuaris, augmentant el seu poder de negociació i oferint capacitat i serveis de tarifes reduïdes per obtenir i captar més quota de mercat.

Els operadors internacionals de transport ofereixen serveis i activitats sobre una base global i controlen les terminals. Posseeixen també un gran poder de negociació i són molt competitius. En la actualitat dirigeixen les seves estratègies cap al control dels nous segments de la cadena de subministraments (tals com ferrocarrils, transport terrestre, serveis *feeders*, terminals interiors i zones logístiques).

Dins d'aquest taulell canviant cal analitzar i tenir en compte el rol de les Autoritats Portuàries, ja que el seu paper està canviant. Les Autoritats Portuàries es distancien cada cop més de les activitats operacionals relacionades amb la provisió de serveis portuaris. Slack (1993), emfatitza que l'entrada de les entitats privades en la provisió dels serveis portuaris i l'augment dels serveis logístics fa que els ports es converteixin en mers peons en l'ampli i complex joc d'interessos globals i de les xarxes intermodals.

Brooks (2004), afirmava que s'escull al transportista abans que al port, doncs aquest es seleccionat per aquell. En aquest sentit, una de les funcions de les

Autoritats Portuàries és la redefinició dels objectius del port, redirigint el procés en cooperació amb altres socis i desenvolupant una xarxa d'actors econòmics i institucionals; en suma, coordinant una xarxa estratègica.

Per tant, el nou rol i el paper de les Autoritats Portuàries és augmentar la eficiència de tot el sistema.

Tenint en compte les definicions, explicacions i arguments anteriors, i el paper canviant de les Autoritats portuàries esmentat, les grans tendències actuals s'expliquen a continuació:

- La primera és que amb el desenvolupament dels ports *landlord* es redueix el control jeràrquic i les companyies adopten decisions que influeixen de manera directa en la competitivitat. Per tant, les companyies privades participen cada vegada més en les activitats operacionals del port.
- La segona tendència és que es procedeix a estendre els conceptes de competitivitat cap al *hinterland*. Bàsicament això significa que les actuacions dels ports estan vinculades al sistema de transport i logística relacionat amb el *hinterland* portuari. D'aquesta manera, les companyies portuàries distribueixen cada cop més els recursos cap a l'interior.
- La tercera de les tendències està relacionada amb la concessionarització també esmentada. Els nous comportaments estratègics empresarials radiquen en la sol·licitud de llicències per nous desenvolupaments i expansions. És a dir, els ports han de demostrar la seva contribució i valor per a la societat, tant en termes econòmics, socials, com ambientals.

A la vista de les tres tendències es poden extreure les següents conclusions:

- Les Autoritats Portuàries tenen cada cop més necessitats de desenvolupar capacitats i recursos per afrontar els nous reptes. Poden fer-ho de forma individual o bé associades a altres.
- Les Autoritats Portuàries estan en una posició institucional adequada per crear i estimular accions col·lectives.
- Les Autoritats Portuàries participen cada vegada més i de manera més intensa en el funcionament i desenvolupament dels *hinterlands*. Els usuaris del port cada vegada tenen més en compte la actuació i els serveis oferts al llarg de tota la cadena a l'hora de seleccionar el seu port d'ús.
- Les Autoritats Portuàries han de reforçar les seves capacitats d'arrossegament tant per les activitats dins dels límits portuaris com de les activitats en el propi *hinterland*.

- Sistema portuari espanyol i regulació europea

Ara que ja es té una visió general del sistema portuari, ens centrarem amb les característiques del sistema portuari espanyol. Primerament es començarà per

una breu introducció a com es troba actualment el procés de regulació dins d'Europa, ja que posteriorment el sistema espanyol es trobarà englobat dins d'aquest marc.

Tot i la gran importància dels ports en l'economia dels Estats membres de la Unió Econòmica, la regulació no va ser inclosa en el Tractat constitutiu de la Comunitat Econòmica Europea (Tractat de Roma, 1957). La Comissió de la Unió Europea va redactar un document el 1992 on s'indicaven els canvis més significatius en política comú de transports, identificant la necessitat de concebre un sistema de transport a nivell europeu i establint els elements bàsics de les anomenades xarxes transeuropees.

L'interès i l'esforç de la Unió Europea per establir xarxes europees va quedar reflectit en el Tractat de Maastrich (1992), encara que el plantejament de les infraestructures és encara competència dels diferents Estats membres. Aquell mateix any la Comissió Europea va indicar la necessitat d'incloure els ports en les xarxes transeuropees.

El 1994 la Comissió va aprovar una proposta per establir les directrius per a desenvolupar les xarxes transeuropees. El resultat va ser la creació d'un grup d'experts dins de l'Oficina General del Transport Marítim de la Comissió per tal de redactar aquestes directrius.

El 1997 la Comissió de la Unió Europea va presentar el Llibre Verd dels Ports i les Infraestructures Marítimes que conclou que la regulació a nivell europeu ha de ser desenvolupada amb mires a aconseguir una liberalització sistemàtica dels serveis en els principals ports amb tràfics internacionals. El debat girava en torn a tres aspectes: incloure els ports en les xarxes transeuropees, la desregulació dels serveis portuaris i el finançament públic dels ports i de la infraestructura portuària. La proposta va ser inclosa dins d'una directiva que va ser aprovada l'any 2003 al Parlament Europeu. Aquesta directiva sobre l'accés al mercat dels serveis portuaris és aplicable a tots aquells ports amb un volum de tràfic marítim superior a 1,5M de tones o 200.000 passatgers anuals. S'hi autoritza l'alliberació de serveis portuaris a excepció del pilotatge.

La posició dels diferents Estats membres sobre el paper dels ports en la Unió Europea sempre ha depès de la importància dels ports en els seus respectius sistemes de transport. Ara bé, les reaccions a tal Directiva van ser dispers: mentre alguns països es van oposar com Suècia o el Regne Unit, la majoria li van donar suport, com Espanya.

Alguns països segueixen una política en que totes les despeses són traslladades al consumidor (visió anglosaxona), mentre que altres intenten compatibilitzar tots els beneficis i despeses associats a la regió d'influència del port (visió continental). En aquest últim cas, l'objectiu de la generació de treball es considerat de vital importància.

Dins d'aquest marc europeu s'hi troba doncs la regulació del sistema portuari espanyol, que parteix de la llei 27/1992 sobre els *Puertos del Estado y la Marina mercante*, la qual fou modificada per la llei 62/1992. La que actualment està

en vigor és la llei 33/2010 introduïda amb el Real Decret Legislatiu 2/2011 que modifica l'anterior llei 48/2003 de Règim econòmic i de prestació de serveis als ports d'interès general, que essencialment ha suposat una major autonomia dels ports i un impuls a la gestió comercial dels ports, en la línia també d'aquesta anterior.

Al sistema portuari espanyol s'hi poden diferenciar dos tipus de ports: els d'interès general, propietat de l'Estat, i els que no ho són (pesquers i/o esportius), propietat de les Comunitats Autònomes. Els primers són aquells involucrats en el comerç marítim internacional i la seva zona d'influència abasta més d'una Comunitat Autònoma.

En les lleis també es fa referència al model de gestió i organització dels ports d'interès general. A l'Autoritat Portuària se li confereix la responsabilitat de l'organització i la gestió del port i per a això se li dota d'autonomia per a gestionar el pressupost del qual disposa, encara que s'haurà de coordinar i sotmetre al control de l'Organisme Públic *Puertos del Estado*. Aquest, recull tots els recursos de tot el sistema de ports i forma un fons de compensació per inversions, d'aquesta manera es redueixen les necessitats de subvencions o de fons de la resta del pressupost de l'Estat. Així, amb els ingressos percebuts per les Autoritats Portuàries s'ha d'aconseguir tant l'equilibri econòmic i financer de tot el sistema portuari com també de cada port en concret.

La gestió de les 28 Autoritats Portuàries i els 46 ports d'interès general ha d'obeir a criteris d'eficiència, economia, productivitat i seguretat.

Els serveis de manipulació de carga tenen una regulació específica. Es tracta d'un servei públic que està sota la responsabilitat de l'Estat, aquest es propietari de la Societat Estatal de l'Estiba i la Desestiba (SEED) en la qual hi són presents els ports d'interès general. En aquest decret es permet l'entrada de les empreses a les activitats portuàries de càrrega i descàrrega mitjançant contractes administratius. Totes les firmes que ofereixen aquest servei han de participar en el capital social de la SEED, encara que més del cinquanta per cent d'aquest és de l'Estat, a efectes de mantenir el control de la societat, Així mateix, les diferents Autoritats Portuàries han de fixar els màxims preus que han de cobrar les empreses d'estiba i desestiba pels serveis de càrrega i descàrrega. Aquesta regulació ha variat en alguns punts recentment amb l'entrada del Real Decret-Llei 8/2017, amb canvis pel que fa referència en major part a l'actual règim de gestió dels treballadors d'aquest àmbit.

El sistema portuari espanyol es caracteritza per posseir una àmplia tipologia de recintes, una possible classificació seria la següent:

- En funció del seu volum de tràfic i de la seva especialització.
- Per la seva inserció a les xarxes marítimes intermodals, ja siguin línies regulars com de transbord.
- Existeixen ports amb una forta i notable vinculació als *hinterland* més pròxims, i, amb això, responen als fluxos de importació i exportació.

- És notable l'amplia dispersió dels ports al llarg del perímetre coster espanyol, de manera que cap espai econòmic espanyol queda sense estar connectat a un port.

Podríem dir que en suma el sistema portuari espanyol està distribuït en quatre zones: la cantàbrica, la gallega, la sud-meridional que inclou les mediterrànies i les sud-atlàntiques, i finalment els ports de Canàries.

A més, el sistema portuari espanyol presenta un alt grau d'especialització i diversificació. N'hi ha especialitzats en grànells líquids, altres en grànells sòlids, en mercaderies, en moviment de contenidors, i en tràfic de creuers.

El sistema té una ampla capacitat per poder respondre a nous increments de tràfics i a les noves línies que escalin en les xarxes portuàries.

La crisi econòmica va marcar un fort impacte en les transaccions comercials, tal com veiem en la caiguda de tràfics portuaris del bienni 2008-2009. Com a resposta a aquesta situació, es van posar en marxa nous postulats recollits a la *Ley de Puertos* 33/2010 a la que ja hem fet referència anteriorment, que apostava clarament per la flexibilitat tarifària, per majors compromisos relacionats amb l'eficiència i la competitivitat tant intra com inter portuària, i per una major participació privada en els serveis tècnic-nàutics i en les concessions de terminals. Els anys 2010 i 2011 presenten una millora a nivell de tràfics i per tant milloren econòmicament i financerament.

Pel que fa als ingressos, estan formats essencialment per les taxes d'utilització especial de les instal·lacions portuàries (54%), i més específicament les taxes de mercaderies (26%) i al vaixell (19%). També destaquen els ingressos procedents de les concessions i de les autoritzacions que es liquiden a través de les taxes d'ocupació privativa del domini públic portuari i de les taxes d'aprofitament especial del domini públic (38%). D'aquesta manera, el 83% del total d'ingressos el componen les taxes del vaixell, de la mercaderia i de les concessions.

Pel costat de les despeses les principals partides són les de personal (29%), les d'explotació (33%) i les amortitzacions de l'immobilitzat (38%).

Les 10 principals Autoritats Portuàries per *Importe de Cifra de Negocios* (ICN) són Barcelona, Valencia, Algeciras, Bilbao, Balears, Les Palmes, Tarragona, Huelva, Gijón i Tenerife, que sumen el 70% del total. Així mateix, els ports amb major nivell d'ingressos i amb majors volums de tràfic són els que millor ratis presenten pel que respecta al volum de fluxos de mercaderies.

Si analitzem les dades d'eficiència de les terminals i dels ports en relació a les seves infraestructures i equipaments, es veu que les Autoritats Portuàries d'Algeciras, Ferrol, Cartagena, Balears, Avilés i València són les que obtenen els millors resultats per euro invertit i les que presenten millors ratis en termes de contenció de despeses i maximització dels beneficis

A tot això i igual que en la tendència general, els ports petits són més vulnerables a les pràctiques limitadores de la competència degut al volum de tràfic portuari.

En definitiva, el sistema portuari espanyol es sustenta en un esquema que permet combinar propietat pública de les infraestructures portuàries com les obres d'abric, amb la propietat privada de la superestructura, per exemple els equips de manipulació. La Autoritat Portuària és qui determina les condicions sota les quals la iniciativa privada ha d'operar.

3. TÈCNIQUES DE PREDICCIÓ BASEDES EN SÈRIES TEMPORALS

3.1. INTRODUCCIÓ ALS MODELS DE PREDICCIÓ

Dins de la predicció del tràfic portuari ha predominat la utilització de dos metodologies tradicionals diferenciades en la predicció, una basada en el PIB i una altra basada en la recerca i processament de les dades referents a l'entorn portuari, que és tant teòrica com de camp.

- Aproximació pel PIB

Es basa amb l'assumpció que els indicadors econòmics poden determinar la demanda de càrrega. Aquesta hipòtesi pot ser certa a la pràctica i en aquest cas es simplifica utilitzant simplement el PIB (Indra, Noteboom, Parola, Satta, Persico, 2015).

Aquest mètode funciona amb multiplicadors adoptats segons la variació de l'indicador. Aquest mètode es pot combinar amb altres mètodes qualitius de predicció que més endavant es presentaran.

No sempre dóna garanties de funcionament ja que depenent del tipus de port al que s'apliqui igual no existeix una relació directa entre l'augment del PIB i el tràfic portuari, és el que es coneix com la dicotomia entre el nombre de contenidors i el contingut del contenidor. Pot passar que el PIB estigui elevat perquè un sector va especialment bé, però el sector al qual es dediqui majoritàriament el port estigui en hores baixes, el PIB creixerà però no es correspondrà amb la realitat del port.

Un segon aspecte a tenir en compte en aquest mètode és el fet d'utilitzar relacions fixes en els multiplicadors en relació al PIB, que podria donar prediccions que diferirien cada cop més de la realitat.

Podríem dir que bàsicament el mètode ens serviria per detectar punts de canvi en la dinàmica en el tràfic portuari.

- Recerca teòrica i de camp

Fa referència a la interacció d'individus i d'organitzacions dins del sector industrial que posen en comú la seva visió i els seus punts de vista. Pot ser un bon mètode sempre i quant els entrevistats no tinguin una visió molt estreta del mercat, confiiïn només amb un tipus concret de predicció o que incloguin variables externes que en realitat siguin poc rellevants (Indra et al, 2015).

És considerat un bon mètode a curt termini, però que perd eficàcia a mig i llarg termini. No obstant, aquest mètode pot ser bo des del punt de vista de complementar l'anàlisi estadístic, en el sentit que aquesta recerca ens pot aportar informació en relació al punt clau que afecta el transport en un sistema portuari determinat. Per exemple una anàlisi per sèries temporals també ens pot complementar l'anàlisi econòmic anterior, o també com a exemple aquesta

recerca ens pot permetre conèixer millor l'estructura de mercat en el qual ens movem per prendre millors decisions de cara al futur, com poden ser el govern i els canvis en la política i tecnologies, que podrien tenir un impacte significant.

3.2. TÈCNIQUES GENERALS DE PREDICCIÓ

La predicció és una eina molt útil i necessària a l'hora de prendre decisions a curt, mig i llarg termini. A continuació, veurem un resum de les metodologies més utilitzades.

Podem diferenciar bàsicament tres tipus de tècniques de predicció: el primer els models qualitatius, el segon l'anàlisi de series temporals i el tercer els models causals, aquests dos últims els englobem dins del que anomenem els models quantitius (Indra et al, 2015).

El primer fa referència als models que utilitzen dades qualitatives, com podria ser l'opinió d'un expert, per exemple, i informació referent a situacions especials que puguin influir a l'hora de prendre una decisió, tenint o no en compte el passat. El segon en canvi es fixa en el model o el patró que s'ha seguit en el passat i els canvis que ha anat patint, per tant, confia enterament de les dades i informació històrica. El tercer en canvi utilitza informació molt refinada i concreta en funció a les relacions entre els elements del sistema, i es prou potent com per detectar i tenir en compte de forma formal els esdeveniments més rellevants durant el període analitzat, és a dir, que en aquest cas també serà important el passat.

Cada tècnica té els seus avantatges i desavantatges, i serviran millor o pitjor en determinades situacions. A més, els models estan sotmesos a l'efecte del temps de planejament, significant que un model pot ser correcte en una determinada situació a curt termini, però pot ser una molt mala decisió si ens marquem la predicció a llarg termini com a fita.

- Models qualitatius de predicció

Aquests models són models de predicció que confien en el criteri expert dels humans combinat amb escales de valoració de determinats criteris, més que no pas en dades que no poden ser mesurades ni verificades amb la màxima seguretat.

L'utilitzarem quan les dades de les que disposem són escasses i ambigües, o quan estem analitzant la implementació de noves tecnologies de les quals encara no en disposem de molta informació.

Aquests models tenen els avantatges de poder incloure en la nostra anàlisi dades no numèriques per completar dades que poden ser ambigües o incompletes, i que ens permet la introducció del criteri i experiència de gent realment experta en el tema tractat, que ens pot aportar una visió més en perspectiva sobre l'assumpte en debat. En canvi, els desavantatges d'aquest model és que no és sempre precís, i que podem caure en la trampa de confiar

massa la predicció a una visió concreta d'algun dels experts i córrer el perill que si aquesta és incorrecta, llavors la predicció també ho serà.

Alguns exemples d'aquests models són: mètode Delphi, predicció basada en el mercat, predicció de consens, predicció visionària, predicció basada en analogia històrica, etc.

- Anàlisi de sèries temporals

Aquesta tècnica és utilitzada quan disposem de dades consistents al llarg dels anys respecte el producte o de la producció que estem analitzant. Amb aquestes dades mesurarem no tan sols els ratis amb la informació que disposem al llarg dels anys, sinó també els ratis de canvi en els canvis que s'hi han produït. Per tant podem dir que els objectius que té aquest model són els següents:

- Descriure les dades observades i obtenir unes mesures simplificades que resumeixin les propietats de les series temporals. D'aquesta manera mirar d'obtenir tendències, estacionalitat, *outliers* i punts de canvi de tendència.
- Identificar el procés de generació de les dades que generen la variable aleatòria de la qual en tenim disponible les seqüències observades dins de la sèrie temporal.
- Predir i fer estimacions sobre possibles valors futurs a partir de les dades històriques. Això està basat en el principi de que el comportament d'un fenomen en el passat es mantindrà també en el futur.

Les dos aproximacions principals per dur a terme aquest procés són el mètodes clàssics d'aproximació a l'anàlisi de series temporals, i els mètodes derivats de l'aproximació de Box-Jenkins, també conegut com anàlisi de series temporals modern (Peng, 2009).

- Models causals

Aquests models són l'eina de predicció més sofisticada a l'hora d'expressar matemàticament les relacions causals més rellevants i representatives. Incorpora els resultats de l'anàlisi de sèries temporals tenint en compte tot el conegut de les dinàmiques dels sistemes de fluxos que s'estiguin analitzant, en aquest cas seria els de transport portuari (Indra et al, 2015).

Tenen les avantatges de que tenen un alt poder explicatiu i permet entendre relacions entre diferents variables, i a més no es necessita l'ús de grans models econòmics ja que els models més simples funcionen bastant bé. No obstant, té els desavantatges següents: d'acord als principis d'econometria les variables amb coeficients insignificants hauran de ser menyspreades; que el canvi en les relacions entre variables augmenta amb el període projectat; i que el seu cost i temps requerit poden ser elevats.

Alguns exemples d'aquests mètodes en poden ser regressions múltiples complexes i models que integrin models econòmics també complexes en conjunció amb indicadors relatius al sector que estem tractant.

Seguint amb la introducció del models i les tècniques generals de predicció, arrodonirem la introducció d'aquest concepte amb alguns principis i riscos de la predicció que ens seran útils en les diferents explicacions posteriors, i ens permetrà aportar una visió més àmplia i objectiva en les valoracions dels models mateixos.

3.3.PRINCIPIS I RISCS DE LA PREDICCIÓ

La predicció estadística o *forecasting* és l'art i la ciència de predir a partir de dades, amb o sense saber d'entrada quina equació s'hauria d'utilitzar. En altres paraules, seria la forma d'intentar saber quina forma o com serà el futur, partint de la base de que serà una cosa similar al nostre present, però una mica més llarg.

Després de fer unes primeres proves i observar el tipus de dades que tenim, la idea és arribar a tenir un model que sembla que porta a prediccions útils i que a més ens aporten informació que en principi no coneixíem.

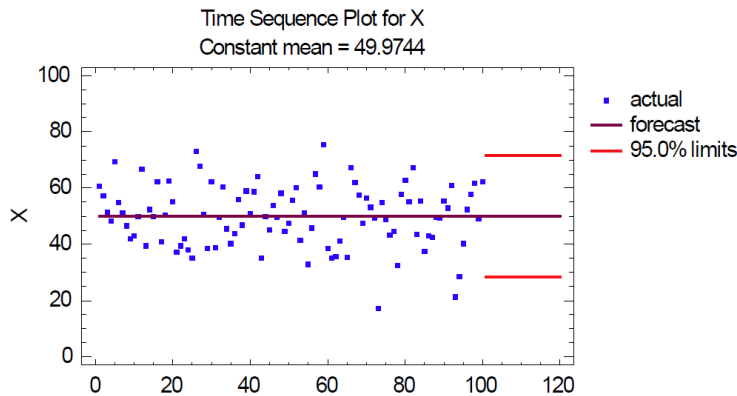
- Senyal vs. soroll

La senyal és l'element o el component que es pot predir, i el soroll és el que deixem enrere, ja que no aporta. La forma tècnica d'entendre el soroll és aquella en que es compleix que el conjunt de dades són "variables aleatòries independents i idènticament distribuïdes".

En la determinació de la senyal, si errem en detectar una que està a la vista o detectem una senyal que no és realment la que hi tenim, correm el risc que totes les prediccions seran, com a molt, poc òptimes i en general portarà a malentesos perillosos de cara als nostres estudis.

- Introducció d'alguns dels casos o patrons més comuns de *forecasting*

El cas més simple són punts que varien al voltant d'una línia horitzontal, com podem veure a continuació:

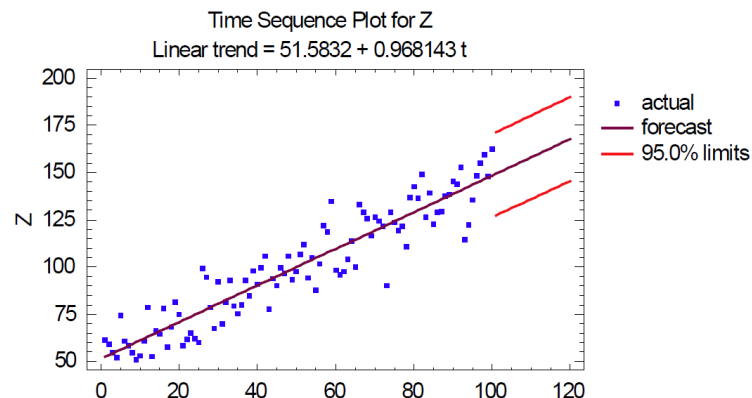


Il·lustració 1 Patró tipus *mean model* (Nau, 2014)

El model de predicció més apropiat per aquest tipus de dades és el *mean model* o model de la mitjana, que és un cas especial de regressió en que existeix un terme que anomenem *intercept* però no variables independents.

Tot i que sembla un model molt simple, no ho és tant com ens podem pensar. Es necessita estimar la mitjana com també la desviació estàndard de les variacions al voltant de la mitja, i la desviació estàndard ha de ser utilitzada adequadament per a calcular els intervals de confiança de les prediccions.

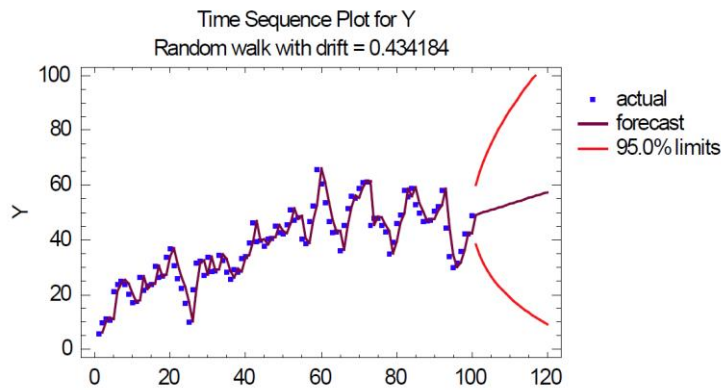
Un altre patró interessant són aquelles variables aleatòries independents i idènticament distribuïdes que varien al voltant d'una línia de pendent en un gràfic de la nostra variable d'interès en relació a una altra variable, que indica un cert grau de correlació entre dos variables, podem veure l'exemple a continuació:



Il·lustració 2 Patró tipus Regressió lineal simple (Nau, 2014)

En aquests casos el model de predicció més apropiat podria ser el model de regressió simple. En el cas que l'eix X fos el temps, aquest model es diu model de tendència lineal.

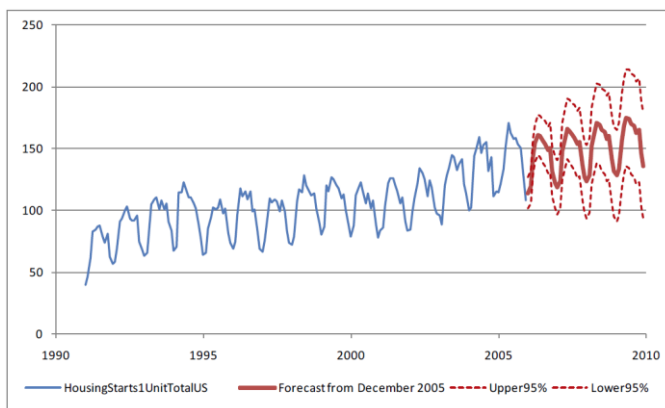
Un model més que també té un patró interessant és aquell en que una sèrie temporal presenta canvis de variables aleatòries independents i idènticament distribuïdes d'un període al següent, com es pot veure en la següent gràfica:



Il·lustració 3 Patró tipus Model del camí aleatori (Nau, 2014)

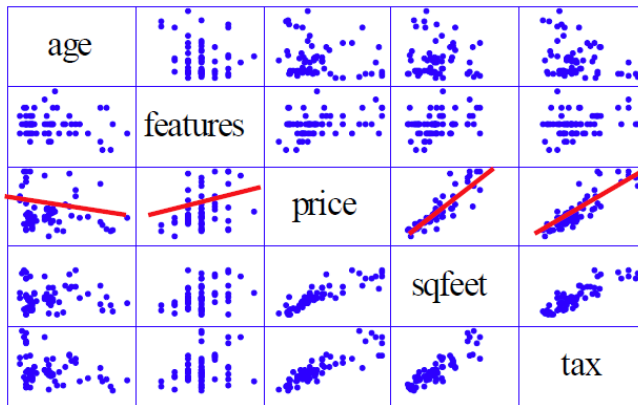
El model de predicció adequat per aquestes series sol ser el *random walk model* o model del camí aleatori, com l'anomenarem en aquesta memòria, que es diu així perquè la variable va donant passades aleatòries cap amunt i cap a baix mentre va avançant. Aquest model és de gran importància en l'anàlisi de dades financeres. En aquests models, encara que directament podem pensar que es tracta d'un *random walk*, hem d'analitzar de forma estadística les propietats d'aquests passos, per veure si realment són aleatoris o si tenen alguna propietat no aleatòria.

Un altre patró que és vist comunament en les dades de les sèries temporals és aquest en que les variables independents i idènticament distribuïdes varien al voltant d'un patró de períodes estacionals com en el cas següent:



Il·lustració 4 Patró tipus Model estacional (Nau, 2014)

Un altre tipus important de patró és aquell de correlacions entre varies i diferents variables, que vindrien a ser com els típics *scatterplots* X-Y però en aquest cas en forma de matriu perquè estem analitzant diverses variables. Un exemple en podria ser el següent:



Il·lustració 6 Patró tipus Model matriu diverses variables (Nau, 2014)

Alguns d'aquests casos simples esmentats seran desenvolupats posteriorment en profunditat durant el transcurs d'aquesta memòria.

- Riscs de la predicció

La predicció no és un assumpte trivial. Es distingeixen principalment tres tipus de fonts dels quals poden proveir la majoria dels errors que podem tenir al fer les nostres prediccions, controlant aquestes fonts podem mirar de reduir-los (Nau, 2014). Són els següents:

1. Risc intrínsec: és la variació aleatòria que existeix més enllà de la utilització i el processament que em fet a les nostres dades amb les diferents eines que tenim a disposició. És a dir, és el soroll del nostre sistema.

Normalment es mesura amb la desviació estàndard estimada de l'error estimant un coeficient o predint un valor d'alguna variable, ens hi referirem simplement com l'error estàndard. Una manera possible de reduir aquest tipus d'error seria refinant el model, això és, agafant més o millors variables explicatives.

2. Risc paramètric: és el risc degut als errors a l'estimar els paràmetres del model de predicció que s'està utilitzant, sempre sota la hipòtesi que s'està utilitzant el model correcte d'acord amb les nostres dades en primer lloc.

Aquest error és normalment molt més petit que el risc intrínsec i el mesurem amb l'error estàndard de l'estimació del coeficient. Una manera de reduir-lo pot ser simplement obtenint una mostra més gran de dades, tot i que en la predicció de series temporals una mostra més gran de dades no sempre és positiu. Amb més dades pot ser que les dades més antigues no siguin representatives amb les condicions actuals i les prediccions surtin desajustades.

L'error estàndard de la predicció que es produeix amb el model és calculat amb una formula que inclou l'error estàndard del model i els

errors estàndards dels coeficients, d'aquesta manera es tenen en compte els dos tipus de riscos, tant l'intrínsec com el paramètric. A més, l'error estàndard de la predicció és sempre més gran que l'error estàndard del model, per tant podríem dir que aquest és el límit inferior de la precisió de la nostra predicció. El gran que sigui aquest error dependrà doncs dels errors estàndards dels diversos coeficients i també de la distància a la que es troben els valors de les variables independents respecte els seus propis valors mitjos.

Com no sempre és possible eliminar els riscos intrínsecs i els riscos paramètrics, és important que ens intentem quantificar uns valors realístics sobre quins poden ser els límits d'aquests valors.

3. Risc del model: és el risc d'escollir un model equivocat, és a dir, error al fer les hipòtesis sobre si el passat realment s'assemblarà al futur i com ho farà.

Normalment aquest és l'error de predicció més freqüent, a més, no disposem d'un error estàndard amb el que mesurar o calcular-lo, ja que cada model s'assumeix a si mateix com a correcte. La manera de reduir-lo o intentar evitar-lo és seguir bones pràctiques estadístiques en tot el desenvolupament de la predicció.

Aquest risc pot arribar a ser molt alt si confiem únicament de les regles més simples o del software automàtic de predicció, sense entendre primer el tipus de dades que tenim, sistemàticament explorant-la, utilitzant el criteri propi i l'experiència, i realitzant diferents tests sobre les hipòtesis que anem assumint al nostre model. No existeix cap fórmula màgica.

Davant de la pregunta sobre quin és el model vàlid, o almenys el que no és òbviament dolent, es necessiten eines per a respondre-la. Una eina senzilla és analitzar si els errors s'assemblen a soroll pur, és a dir, variables aleatòries independents i idènticament distribuïdes. Si no ho són, significa que hi pot haver algun patró al darrere que ens permet reduir-los ajustant al model a través del qual es pot explicar tal patró.

3.4.MÈTODES I MODELS

3.4.1. Estadística bàsica i model de predicció més simple: la mitjana de la mostra

Es considera una certa variable corresponent a una sèrie temporal (un conjunt d'observacions ordenades en el temps), aquesta variable s'anomenarà X . Es suposa que tenim un cert nombre d'observacions a la nostra disposició.

El model de predicció més simple que es pot considerar serà el *mean model* o model de la mitjana, el qual assumeix que la sèrie temporal consisteix en variables aleatòries independents i idènticament distribuïdes. Sota aquesta

hipòtesi, el següent valor hauria de ser predit de manera que fos igual que la mitjana històrica de la mostra si l'objectiu és minimitzar l'error quadrat mitjà.

Revisem algun dels conceptes més bàsics d'estadística per al seu posterior ús en la continuació de l'explicació (Devore, 2016):

X = variable aleatòria, els seus valors individuals seran denotats com x_1, x_2 , etc.
 N = grandària de la població sencera de valors de X (possiblement infinita).
 n = grandària de la mostra finita de X .

La població mitjana ("vertadera") μ és la mitjana de tots els valors de la població.

$$\mu = \sum_{i=1}^N x_i / N \quad (1)$$

La variància de la població σ^2 és la desviació mitjana al quadrat respecte el valor mig vertader:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \quad (2)$$

La desviació estàndard de la població σ és l'arrel quadrada de la variància de la població. En aplicacions com la de predicció no observarem mai tota la població, ja que el nostre problema serà predir a partir d'una mostra finita. Per tant, els estadístics com les els valors mitjos i les desviacions estàndards han de ser estimades amb un cert error.

El valor mitjà de la mostra és la mitjana de tots els valors en la mostra:

$$\bar{X} = \sum_{i=1}^n x_i / n \quad (3)$$

Aquesta serà la predicció puntual del model de la mitjana per tots els futurs valors de la mateixa variable.

La variància de la mostra s^2 és la desviació quadrada mitja respecte la mitjana mostral, excepte pel factor de $n - 1$ a més de la n al denominador:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1} \quad (4)$$

La desviació estàndard de la mostra és l'arrel quadrada de la variància de la mostra, denotada amb la lletra s .

El factor de $n - 1$ al denominador de la variància de la mostra s'utilitza enlloc de la n degut al grau de llibertat de l'error que s'ha utilitzat per calcular la

mitjana mostral de les mateixes dades. Podem dir que l'ajustament correcte per tal que ens quedi un estimador sense biaix de la variància vertadera és dividir la suma de les desviacions quadrades pel nombre de graus de llibertat, no pel nombre de dades que tenim.

Llavors, és necessari entendre per a què és important l'error quadràtic quan fem aquest tipus de càlculs. És comú mesurar la variabilitat en termes de desviacions quadrades mitges enlloc de la desviació mitjana absoluta al voltant d'un valor central, bàsicament degut a que l'error quadràtic té moltes propietats positives per a nosaltres:

- El valor central al voltant del qual la suma de les desviacions quadràtiques són minimitzades és, de fet, la mitjana mostral. Per tant, quan ajustem models de predicció minimitzant la suma dels seus errors quadràtics, estem implícitament calculant mitjanes, fins i tot quan estimem varies coses a la vegada.
- Les variàncies són additives quan les variables aleatòries que són estadísticament independents són afegides de forma conjunta. I com les variàncies i les covariàncies tenen un paper clau en la teoria de la distribució normal i l'anàlisi de regressions, ens pot anar bé aquesta propietat.
- Des d'un punt de vista teòric de les nostres decisions, els errors grans tenen la majoria de les vegades pitjors conseqüències que els petits, per tant l'error quadràtic és més representatiu com més grans siguin aquests.
- Les variàncies i les covariàncies tenen un paper clau en la teoria de la distribució normal i l'anàlisi de regressions.

Seguint amb els conceptes, l'error estàndard de la mitjana és el següent:

$$SE_{mean} = s/\sqrt{n} \quad (5)$$

- Aquesta és la desviació estàndard de l'error que tindrem utilitzant la mitjana mostral \bar{X} com a estimador de la mitja vertadera μ si repetíssim aquest exercici amb altres mostres independents de grandària n .

- Mesura la precisió de l'estimació que fem de la mitja vertadera (que desconeixem) d'una mostra limitada de dades.

- Com més gran sigui la grandària n més petit serà SE_{mean} i la distribució de l'error a l'estimar la mitjana s'aproximarà a la normal.

- En concret i com veiem per la formula (5), 4 vegades més de dades redueix l'error estàndard de la mitjana a la meitat.

Per entendre la relació entre la desviació estàndard i l'error estàndard en relació a les prediccions, es pot dir que l'error estàndard és la desviació estàndard de l'error al estimar o predir alguna cosa.

La mitjana no és l'únic estadístic per a mesurar el valor típic o representatiu d'una població donada. Per exemple la mediana (percentil 50) també és un estadístic de resum que descriu un nombre representatiu d'una població.

Val a dir però, que els models més utilitzats en les prediccions, com els models de regressió, es centren en les mitjanes (junt amb les desviacions estàndards i les correlacions) com la estadística descriptiva clau, i les prediccions puntuals es centren en valors mitjos i no en els valors de la mediana, ja que d'aquesta manera es minimitzen els errors quadràtics mitjos.

Un altre problema que ens podem trobar al fer prediccions amb gran nivell de detall, és que la mediana d'una variable en un sol període podria ser 0, per tant donaria una distribució trivial i que no aportaria informació.

A més, les transformacions no lineals a les dades poden ser utilitzades per transformar les distribucions a simètriques, permetent que les dades puguin ser ajustades de forma correcta amb models que es centren amb valors mitjos.

- Predicció amb el model de la mitjana

Abans de començar, denotem cada terme:

- Notarem amb \hat{x}_{n+1} la predicció de x_{n+1} basada en les dades que tenim fins al període n .
- Si la x_{n+1} és assumida independent respecte la mateixa població que la mostra x_1, \dots, x_n , llavors la predicció que minimitza l'error quadrat mig és simplement la mitjana de la mostra:

$$\hat{x}_{n+1} = \bar{X} \quad (6)$$

De forma especial per aquest model, la desviació estàndard de la mostra (s) és el que anomenem error estàndard del model, és a dir, la desviació estàndard estimada del risc intrínsec (Nau, 2014). Llavors, l'error que podem esperar a l'utilitzar \hat{x}_{n+1} per la predicció de x_{n+1} és el que anomenem l'error estàndard de la predicció (SE_{pred}), i depèn de l'error estàndard del model i de l'error estàndard de la mitjana:

$$SE_{pred} = \sqrt{s^2 + SE_{mean}^2} = s \sqrt{1 + \frac{1}{n}} \approx s(1 + \frac{1}{2n}) \quad (7)$$

- SE_{pred} mesura l'error de la predicció sota la hipòtesi que el model és correcte.
- s l'error estàndard del model mesura el risc intrínsec, és a dir, el soroll estimat en les nostres dades.
- SE_{mean} és l'error estàndard de la mitja i mesura el risc paramètric, és a dir, l'error a l'estimar la senyal de les nostres dades.
- Observem per l'expressió final que pel model mig de la mitjana, SE_{pred} és lleugerament més gran que la desviació estàndard de la mostra.

Es pot observar també que en el cas d'elevat al quadrat cada costat de l'expressió el que obtenim és que la variància estimada de l'error de la predicció és la suma de la variància estimada del soroll i la variància estimada de l'error a l'estimar la mitjana. És a dir, es pot descriure de la següent manera:

$$\begin{aligned} \text{Variància del risc en la predicció} \\ &= \text{variància del risc intrínsec} \\ &+ \text{variància del risc paramètric} \end{aligned} \quad (8)$$

Seguint amb la terminologia, la següent definició és el concepte d'interval de confiança. Una predicció puntual hauria d'anar sempre acompanyada d'un interval de confiança per tal d'indicar la precisió que suposadament es vol aconseguir.

Quan ens referim a un interval de confiança del x%, ens estem referint a que l'interval està calculat d'acord a una regla en que s'obtindrà el valor "vertader" dins de l'interval un x% del temps sota les condicions simulades, sempre assumint que el model és correcte. Per tant, de forma més entenedora significa que hi ha un x% de confiança que les dades de predicció futures caiguin dins l'interval de confiança del x%.

En el cas que la distribució del soroll sigui la normal, llavors l'interval de confiança de la predicció de la mitjana és igual al punt predit més o menys un cert valor d'error estàndard en la nostra predicció, i aquest valor s'anomena "t-valor crític". En forma de fórmula podríem escriure-ho de la forma següent:

$$\begin{aligned} \text{Interval de confiança} \\ &= \text{predicció} \pm (t - \text{valor crític}) \\ &\quad * (\text{error estàndard de la predicció}) \end{aligned} \quad (9)$$

Podem veure el t-valor de les taules corresponents a la distribució. Aquesta distribució és la distribució de la quantitat:

$$\frac{(\bar{X} - \mu)}{SE_{mean}} \quad (10)$$

Que és el nombre d'errors estàndards en que la mitjana mostral es desvia respecte la mitjana vertadera quan la desviació estàndard de la població és desconeguda.

Amb tot, es podrà calcular el límit inferior i superior en cada període que ens anirà conformant el que serà el nostre interval de confiança.

3.4.2. Model del camí aleatori (*Random walk model*)

Un dels models més simples però tot i així més importants en l'anàlisi de series temporals és el model del camí aleatori, o de *Random Walk* en anglès. Aquest model suposa la hipòtesi que en cada període la variable fa un pas aleatori més enllà del seu valor anterior, i aquests passos són independentment i idènticament distribuïts en grandària. És equivalent a dir que la primera diferència de la variable és una sèrie en que el model de la mitjana hauria de ser aplicat (Nau, 2014).

A més aquest model de camí aleatori pot tenir o no una constant d'acord a si la distribució de la grandària dels passos tenen o no una mitjana de zero. Al període "n", el k-pas endavant de la predicció que el model de camí aleatori sense constant et dóna per una variable Y és el següent:

$$\hat{Y}_{n+k} = Y_n \quad (11)$$

Per tant, en altres paraules podem dir que el model prediu que totes les variables futures seran iguals a l'últim valor observat. D'acord amb això, el fet d'extrapolar aquest model a prediccions d'un futur molt distant dóna com a resultat una línia horitzontal, com les prediccions fetes amb el model de la mitjana. És a dir, qualitativament les prediccions puntuals d'aquest model es veuen similars a les del model de la mitjana, a excepció de que aquest model és sempre re-ancorat a l'últim valor observat i no a la mitjana de les dades històriques.

Per un model de camí aleatori amb constant, el k-pas endavant de la predicció per un període "n" és:

$$\hat{Y}_{n+k} = Y_n + k\hat{d} \quad (12)$$

On el segon terme de la dreta és la constant que es refereix a l'increment mig entre un període i el següent. Per tant, per prediccions d'un futur molt distant,

aquest model ens donarà una tendència lineal de l'estil d'una regressió lineal de pendent " \hat{d} ", però també és sempre re-ancorat a l'últim valor observat i no ancorada al centre de massa de les dades com en el cas de la regressió lineal.

En el cas del model de camí aleatori amb constant, aquesta pot ser difícil de determinar, i la seva estimació sol fer-se amb els canvis mitjos observats en els diferents canvis de període del passat, que no és més que la diferència entre el primer i l'últim valor en la sèrie dividit per $n-1$:

$$\hat{d} = \frac{Y_n - Y_1}{n - 1} \quad (13)$$

Aquesta serà la pendent entre el primer i l'últim valor de les diferents dades, no la pendent d'una línia de tendència ajustada a les dades per una regressió simple. Aquest mètode per estimar la constant és molt sensible a l'elecció del període històric que utilitzem en l'ajust del model.

En el cas del model de camí aleatori, els intervals de confiança (ja sigui amb els models amb constant o no) es fan molt amples a mesura que la predicció es fa sobre un futur cada cop més llunyà, la qual cosa també el diferencia del model de la mitjana en que els intervals de confiança es mantenen constants amb el temps.

Existeix l'anomenada "regla de l'arrel quadrada del temps", que ve a significar que l'interval de confiança per k -períodes endavant en la predicció amb models de camí aleatori és més ampli que un període de predicció endavant amb una relació igual a l'arrel quadrada de " k " (Nau, 2014). De manera que per exemple una predicció amb 4 passos cap al futur és el doble d'ampla que una d'un sol pas.

El motiu pel qual aquesta propietat és important es degut a que la desviació estàndard de la distància recorreguda en " k " passos no és més que l'arrel quadrada de la variància, per tant l'error estàndard del k -pas és ampliat en un factor de l'arrel quadrada de " k " i no simplement en un factor de " k ".

Amb tot, veiem que l'important és la determinació de l'error estàndard per una predicció d'un pas, i a partir d'aquí es calcula la resta amb la regla de l'arrel quadrada de " k ".

Per determinar l'error estàndard del primer pas en el model de camí aleatori amb constant, ho calculem sabent que aquest és igual a la desviació estàndard de les series diferenciades. Mentre que en el cas del model sense constant serà igual a l'arrel quadrada mitja de les series diferenciades, degut a que el valor mig és assumit com a zero enlloc de ser estimat a partir de la mostra.

Una altra cosa a tenir en compte amb els càlculs a nivell teòric, és que el t -valor de la distribució per calcular l'interval de confiança basada en la predicció i l'error estàndard no és el mateix (Devore, 2016): per al model sense constant, el t -valor és el basat en $n-1$ graus de llibertat, sent " n " la grandària de la mostra,

mentre que pel model amb constant el t-valor estarà basat en $n-2$ graus de llibertat.

Una vegada entrats en matèria i utilitzant el model, un dels estadístics de contrast per comprovar la hipòtesi de que la mitjana sigui significativament diferent de zero és el rati de la mitjana respecte el seu propi error estàndard. Amb aquest simple rati podem identificar si realment és necessari utilitzar la constant en el model per prediccions a curt termini. Aquest estadístic no ens servirà per a prendre una decisió immediata sobre la predicció però si com a eina per a simplificar el model el màxim possible. S'ha de recordar que el fet que sigui sense constant significa que el model tindrà la mitjana en el zero.

A l'hora de treure conclusions a partir del model cal recordar l'esmentat sobre la importància del període de dades escollit. Els resultats que obtinguem dependran en gran mesura de les hipòtesis prèvies, i una de les assumpcions claus és la quantitat de dades passades són considerades com a rellevants.

- Utilitat dels models de camí aleatori

Aquest model es basa en el principi de que les prediccions futures seguiran el patró següent: el pròxim període serà igual que l'anterior més un creixement mig.

Molts models d'anàlisi de sèries temporals molt sofisticades no són més que variacions del model de camí aleatori per a refinar certs punts del model. No sempre el més simple és el millor però és un motiu suficient com per tenir-lo en compte.

D'aquesta manera és un bon model per comparar respecte altres models que estiguem utilitzant, com ara el cas dels models de regressions. Un model de regressió mal escollit segurament funcionarà pitjor que un model de camí aleatori per a les sèries temporals, per molt alta que sigui la seva "R" al quadrat.

Per tant, és un model a tenir en compte tot i no ser el de principal ús en les sèries temporals que estiguem analitzant, ja sigui per comparar amb el model utilitzat, com per poder treure conclusions sobre el model en cas de no tenir-les totes.

3.4.3. Prediccions amb mitjanes mòbils (*Forecasting with moving averages*)

3.4.3.1. Mitjanes mòbils simples (*Simple moving averages*)

Fins al moment s'han vist dos dels models de predicció, el model de la mitjana i el model del camí aleatori. Es pot dir que ambdós representen els dos extrems pel que fa a la predicció a partir de sèries temporals. Un es basa en la hipòtesi que el millor per predir el que passarà demà és la mitjana del que ha vingut passant fins ara, mentre que l'altre assumeix que el millor per predir el que

passarà demà és el que ha passat avui, i la resta de les dades històriques poden ser ignorades.

És dins d'aquest gran ventall de possibilitats on ens mourem dins d'aquest model. El model de mitjanes mòbils parteix de la idea d'agafar una mitjana del que ha estat passant en un cert espai de temps del passat recent, és a dir, busca beneficiar-se de les dos postures.

Moltes sèries temporals de les que ens podem trobar semblaran ser localment estacionàries en certs punts en el sentit que tindran variacions respecte una mitjana local que anirà canviant gradualment amb el temps en una manera no sistemàtica, donant lloc per exemple a mitjanes locals que segueixen un patró cíclic. Si sol ens fiem de la mitjana global de les dades, el model de la mitjana no tindrà en compte aquest patró i el resultat que predirem tendirà a ser sobredimensionat o infradimensionat per a períodes consecutius, depenent de la naturalesa de les dades.

Aquesta tendència pot ser identificada en termes estadístics en els gràfics d'autocorrelació dels residuals (errors). D'aquesta manera podrem observar si l'autocorrelació del patró és positiva o negativa, i conjuntament amb la informació disposada dels intervals de confiança podrem fer una valoració més profunda de si el model o els models utilitzats són els més compatibles amb les dades que tenim. Així mateix, un model que ens doni fora dels intervals tant en predicció puntual com en intervals de confiança haurà de ser descartat ja que alguna de les seves hipòtesis probablement sigui incorrecta.

En el cas de provar el model de la mitjana i el model del camí aleatori i que ni els límits ni els valors dels errors siguin acceptables, podem passar a plantejar-nos utilitzar un altre model com el de mitjana mòbil per tal de predir millor que el model de la mitjana els patrons cíclics, i de no ser tan sensible a canvis aleatoris en les dades entre un període i el següent com ho pot ser el model de camí aleatori.

El model de mitjanes mòbils pot ser calculat de diverses maneres. El cas més lògic és el de fer la mitjana dels "m" valors més recents, donant lloc al que seria el Model de Mitjanes Mòbils Simples (MMS), on l'equació per predir un cert valor de "Y" en un cert període de temps t+1 donades les dades fins a "t" és la següent:

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-m+1}}{m} \quad (14)$$

De la mateixa equació podem veure que el model de camí aleatori no és més que aquest mateix model pel cas de m=1.

El model de les Mitjanes Mòbils Simples té les següents propietats característiques:

- Cada una de les "m" observacions passades té el mateix pes de 1/m en l'equació, per tant, com més gran sigui "m" menys pes tindran les diferents observacions del passat recent. Això implica que a major valor de "m", més "soroll" podrem treure entre període i període, donant lloc a prediccions molt fines a nivell gràfic.
- El primer terme de la mitjana de l'equació es refereix a "1 període d'antiguitat" en relació al punt del temps en el qual començarem a calcular la nostra predicció, el segon període serà el segon període d'antiguitat, i així fins al període "m" d'antiguitat. D'acord amb això, l'antiguitat mitjana de les dades en la predicció serà $(m+1)/2$. Aquesta és la magnitud amb la qual les nostres prediccions tendiran a retardar-se en el seguiment de la tendència o en respondre als *turning points* o punts de canvi de tendència. Per exemple si $m=5$, l'antiguitat mitjana serà 3 i es correspondrà amb el nombre de períodes pel qual la nostra predicció es retardarà respecte el que està passant ara.

D'acord amb aquestes característiques, veiem que existirà un *trade-off* en l'elecció de "m" per tal de no accentuar excessivament alguns d'aquests efectes: escollir entre treure més "soroll" o respondre de forma lenta a les tendències o als canvis de tendència. Si la nostra "m" és petita, el nostre model respondrà relativament ràpid al canvi de tendències, però pel contrari no obtindrem un gràfic estadísticament elegant en el sentit que l'espectre de valors que seguirà agafant el model serà gran. I, pel contrari, si la nostra "m" escollida és gran, el model serà elegant en relació a l'ajustament de les dades, però la reacció als canvis serà excessivament retardada i, per tant, ens pot portar a fer assumpcions a partir del model que no es corresponguin amb la realitat.

- Comparació errors en la predicció entre models

D'acord amb el raonament anterior, necessitem eines per determinar quin és el model correcte i que pugui ser aplicable per prendre decisions al context en el qual ens trobem, o almenys quin és el que funciona millor amb les nostres dades.

Una de les eines en les que ens recolzem per a tals determinacions, serà el càlcul de diferents valors dels errors que s'esmenten a continuació:

- **RMSE (root mean squared error):** el més comú a l'hora de valorar la bondat de l'ajustament, penalitza els errors grans. Aproximadament equival a la desviació estàndard dels errors si la mitjana de l'error fos propera a zero.
- **MAE (mean absolute error):** és la mitjana dels valors absoluts dels errors, no penalitza tant els errors grans perquè en aquest cas no es té en compte el quadrat.

- **MAPE (mean absolute percentage error):** en aquest cas és mirar-se els errors des del punt de vista percentual, que pot ser d'ajuda per estudiar creixements o variacions en les nostres dades.
- **ME (mean error):** indica si les prediccions tenen el biaix alt o baix, hauria de ser proper a zero.
- **MPE (mean percentage error):** mesura el mateix que el ME però en aquest cas en termes percentuals.

Dins d'aquest grup d'errors, generalment el que més ens interessarà serà l'RMSE, ja que normalment ens preocuparà més de cara als càlculs un error molt gran en les nostres aproximacions que errors petits que a priori ens seran poc significants. Normalment els diferents softwares intenten estimar els paràmetres del model pel mètode dels mínims quadrats per tal de minimitzar aquest error. No obstant, tot i que també s'analitzaran altres errors, ens centrarem en aquest estudi tant en l'RMSE, com en l'error absolut mig i en l'error absolut mig en forma percentual, per tal de tenir una visió més general i no centrar-nos en un sol error.

Seguint amb els models, el model de Mitjanes Mòbils Simple pot ser modificat de varies maneres per tant d'ajustar encara més els resultats. En el cas de que hi hagi una tendència clara en les nostres dades, les prediccions amb MMS tindran biaix degut a que el model no conté ninguna constant que capturi aquesta tendència. Per tal de solucionar-ho, afegim una constant de tendència a l'equació de l'MMS, de forma similar a com ho fèiem als models de camí aleatori amb constant:

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-m+1}}{m} + d \quad (15)$$

3.4.3.2. Ajustament exponencial simple (*Simple exponential smoothing*)

El fet de posar pesos iguals en les últimes "m" observacions i cap pes per les observacions anteriors pot ser que no sigui la millor manera de fer les mitjanes per a valors que arriben consecutivament en el temps. D'aquesta forma, una altra manera d'entendre-ho és pensar que els valors passats tenen alguna rellevància, tot i que no la mateixa que les més recents, ja que la hipòtesi de partida és que aquestes seran més importants a l'hora de predir que passarà a continuació. Per tant, per tal d'incloure'ls, el que podem fer es reduir de forma gradual el pes dels valors com més ens allunyem del període més actual de dades.

Una implicació directa és que les prediccions utilitzant aquest mètode seran cada cop menys precises com més ens allunyem cap al futur, és a dir, intervals de confiança més grans.

Així doncs, per tal de pal·liar aquestes deficiències del MMS, el que podem fer és utilitzar aquest model d'Ajustament Exponencial Simple (AES), també conegut com model de Mitjanes Mòbils Exponencialment Ajustades. És molt utilitzat degut al fet que funciona bé i computacionalment és bastant simple.

Hi ha varies maneres d'escriure l'equació de predicció utilitzant el model AES. Una forma és definir unes series "L" que representin el "nivell" (entenent per nivell el valor de la mitjana local) de les series estimades de les nostres dades fins al present. Aquest valor de "L" per un temps "t" és calculat a partir del seu valor previ de la següent forma:

$$L_t = \alpha Y_t + (1 - \alpha)L_{t-1} \quad (16)$$

On α és la constant d'ajustament amb valors entre 0 i 1.

Per tant, el nivell estimat al temps "t" es calcula interpolant entre el valor que acabem d'observar i el nivell estimat prèviament, amb els pesos α i $1 - \alpha$ respectivament. Les sèries de "L" seran més fines gràficament com més propera estigui α de zero, ja que els canvis en cada nova observació de "Y" no seran tan ràpids. També observem que en el model s'assumeix que no hi ha tendència, de manera que no prediu canvi entre un nivell d'un període amb el del següent. Per tant, amb tot això, la predicció per un període t+1 serà simplement el nivell estimat de les series de "L" al temps "t":

$$\hat{Y}_{t+1} = L_t \quad (17)$$

Entenent això, veiem que de forma equivalent el model diu que la predicció del període següent és calculada interpolant entre l'últim valor observat i la predicció que se n'havia fet d'aquest:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t \quad (18)$$

Es pot observar que el cas amb $\alpha = 1$ es correspon amb el model de camí aleatori, i el cas amb $\alpha = 0$ es correspon amb el model de la mitjana, per tant es pot dir que el model AES és una interpolació entre els dos en la forma com aquests dos models responen a les noves dades.

Una altra manera diferent d'escriure l'equació del model AES és a partir de l'error de predicció previ. L'error fet en un període "t" és definit com el valor actual menys el predit, és a dir:

$$e_t = Y_t - \hat{Y}_t \quad (19)$$

Llavors, l'equació de la predicció per un període $t+1$ amb el model AES pot ser expressada com:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha e_t \quad (20)$$

Per tant, l'última predicció és ajustada en la direcció de l'error. Aquesta versió permet una interpretació més bona del valor α , ja que en aquest cas és la fracció de l'error de predicció que es creu que es degut a un canvi inesperat en el nivell de la sèrie, enlloc d'un esdeveniment puntual inesperat.

En el límit quan α tendeix cap a 1, que com hem vist anteriorment es correspondria amb el model de camí aleatori, tota la variació entre un període i el següent seria deguda al canvi en el nivell fonamental enlloc de desviacions temporals.

En el límit quan α tendeix a 0, que com hem vist anteriorment es correspondria amb el model de la mitjana, el nivell fonamental de la sèrie és assumit com a constant i tota la variació entre períodes és atribuïda a desviacions temporals respecte aquest.

Per últim, una altra versió de l'equació de predicció del model AES és la escrita simplement com una mitjana mòbil ajustada exponencialment de tots els valors passats:

$$\hat{Y}_{t+1} = \alpha[Y_t + (1 - \alpha)Y_{t-1} + (1 - \alpha)^2Y_{t-2} + (1 - \alpha)^3Y_{t-3} + \dots] \quad (21)$$

Per obtenir aquesta versió del model, simplement hauríem de partir de la primera versió de l'equació i anar substituint els valors de les prediccions a cada període.

Per tant, un cop vistes les diferents versions i entès el model, es pot senyalar la principal diferència entre aquest model i el model MMS: la predicció amb AES utilitza tots els valors passats però descompta els diferents pesos en un factor $1 - \alpha$ per període, mentre que el model MMS utilitza únicament els últims "m" valors i els dona un pes equivalent d' $1/m$ a tots.

Recordant el concepte d'antiguitat mitjana del model MMS, es pot relacionar amb el valor α del model AES. En aquest cas, l'antiguitat mitjana la podem trobar donant un pes de forma exponencial a les antiguitats dels valors previs de "Y" que apareixen a l'equació. D'aquesta manera per un període d'antiguitat

el pes serà α , per a dos períodes d'antiguitat el valor serà $\alpha(1 - \alpha)$, per tres períodes $\alpha(1 - \alpha)^2$, i així successivament, ens queda:

$$\alpha + 2\alpha(1 - \alpha) + 3\alpha(1 - \alpha)^2 + 4\alpha(1 - \alpha)^3 + \dots = 1/\alpha \quad (22)$$

Obtenim el resultat anterior del càlcul del sumatori a l'infinit.

Per tant, amb això veiem que:

$$\frac{m + 1}{2} = 1/\alpha \quad (23)$$

D'aquesta manera en aquest model l'ajust dependrà del valor α , el qual podem estimar per la relació que guarda amb "m".

Els valors amb una α petita donaran lloc a models més fins gràficament però que tardaran en respondre als canvis de tendència, tal com passava amb els valors grans de "m". Per a valors grans de α el model tindrà resposta més ràpida als canvis de tendència però serà menys fi, també com passava per m petites. Per a valors grans de α , a més, la manera de predir el que passarà també serà més incerta que per a valors petits, com també en seran més amplis els intervals de confiança.

Arribats a aquest punt ens podem fer també el plantejament sobre com podem actuar si les nostres dades tenen una tendència marcada. Com el model AES assumeix que no n'hi ha, el que podem fer com en altres ocasions és afegir una constant que incorpori aquesta tendència al model:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha e_t + d \quad (24)$$

On "d" indica el creixement mitjà per període a llarg termini. Tot i que ho analitzarem més endavant, aquest model concret es correspon a l'ARIMA(0,1,1) amb constant.

3.4.4. Models de regressió

De forma molt simple, es pot definir l'anàlisi per regressions com l'art i la ciència d'ajustar les dades que ens són proporcionades amb línies rectes (Nau, 2014).

En un model de regressió lineal, la variable d'interès o variable dependent és predita a partir de "k" altres variables o variables independents utilitzant una equació lineal. Si "Y" es correspon amb la variable dependent i les diferents "X"

a les independents, la hipòtesi és que el valor de la "Y" al temps "t" en les nostres dades és determinada per l'equació linear següent:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t \quad (25)$$

On les betes són constants i les èpsilons són variables aleatòries incorrelacionades i idènticament distribuïdes normalment amb mitjana zero, que equival al que es correspondria amb el soroll en el sistema. β_0 és també anomenat l'*intercept* del model (valor que prendria "Y" quan totes les "X" són zero) i les diferents β_i són els coeficients multiplicadors de les variables X_i . Les betes, juntament amb la mitjana i la desviació estàndard formen el que anomenarem els paràmetres del model. L'equació corresponent per la predicció de Y_t a partir dels valors corresponent de les "X" és la següent:

$$\hat{Y}_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + \dots + b_k X_{kt} \quad (26)$$

On les diferents "b" són estimacions de les betes i es poden obtenir per mínims quadrats, és a dir, minimitzant l'error quadràtic de la predicció a partir de la mostra.

Aquest és un dels models més simples per predir una variable a partir d'unes altres, i està basat en les hipòtesis següents:

- El valor esperat de "Y" és una funció lineal de les diferents variables "X". Per tant, si una variable X_i varia ΔX_i , "Y" canviarà en una magnitud proporcional de $\beta_i \Delta X_i$. El valor de les diferents betes és sempre el mateix i pot ser un nombre tant positiu com negatiu. Veiem doncs que l'efecte que tenen les diferents variables independents "X" sobre "Y" no és més que la suma dels efectes que cada una de les variables fa per separat.
- Les variacions inexplicades de "Y" són variables aleatòries incorrelacionades, i més en particular les anomenarem no autocorrelacionades en anàlisi de sèries temporals.
- Totes tenen la mateixa variància, és a dir, existeix homocedastitat.
- Totes són normalment distribuïdes.

Tot i que saber utilitzar el model és important, un dels factors que més influència pot arribar a tenir és la recol·lecció de les dades respecte al problema que ens adrecem. L'elecció d'un bon model de regressió depèn en gran mesura de la utilització de dades de les quals en coneixem una procedència fiable. Junt amb això, serà també molt important fer una anàlisi descriptiva prèvia per entendre una mica el patró de les nostres dades i trobar possibles

problemes a temps. Pot ser necessària també una transformació de les nostres dades si hi trobem moltes relacions no lineals o soroll que no segueix una distribució normal o directament no és dependent del temps. Fent això i acabant d'ajustar el model, tan sols quedarà comprovar si realment el model ha respost bé per tal de plantejar possibles altres models a utilitzar o quedar-nos amb el nostre si la precisió es bona.

- Correlació i *regression-to-mediocrity*

El concepte de *regression-to-mediocrity* pot servir per introduir o per acabar d'entendre les relacions lineals que hi podem tenir en el cas de les regressions i com poder fer comprovacions de que ens mantenim dins dels límits que ens esperem en les nostres anàlisis.

El significat del concepte en si, significa que si esperes o saps que t'has de moure per una mitjana aproximada en un determinat aspecte, i obtens de forma successiva resultats molt positius o molt negatius, podríem dir que pots esperar que en períodes següents aquests resultats es reverteixin per a seguir estant dins de la mitjana.

Un exemple fàcil d'entendre aquest concepte és per exemple en els jugadors de bàsquet quan diem que tenen el canell calent perquè els entra tots els llançaments, llavors és fàcilment esperable que en els següents llançaments erri varies vegades per tal de tornar a equilibrar la seva mitjana.

Partint de la idea de fons d'aquest concepte, podem relacionar les nostres prediccions amb gràfics X-Y de dispersió que ens indicaran en una línia de regressió un eix de simetria en la predicció de la distribució conjunta.

Estandarditzant els valors de "Y" i "X" convertint-los en unitats de desviació estàndard respecte la mitjana, la distribució dels punts descriuran aproximadament una dispersió al voltant d'una simetria de 45 graus d'inclinació dibuixats des de l'origen, punt des del qual també partirà la línia de regressió per predir els valors de "Y" respecte "X", tot i que aquesta estarà una mica per sota dels 45 graus de pendent, és a dir, farà una regressió cap a l'eix "x" (Devore, 2016).

Per tant, les variables estandarditzades de "X" i "Y" observades en un període "t" es poden definir de forma simple de la manera següent:

$$X_t^* = \frac{(X_t - \text{mitjana}(X))}{\text{desviació estàndard de la població}(X)} \quad (27)$$

$$Y_t^* = \frac{(Y_t - \text{mitjana}(Y))}{\text{desviació estàndard de la població}(Y)} \quad (28)$$

A més, matemàticament, el coeficient de correlació entre "X" i "Y" és la mitjana del producte dels seus valors estandarditzats:

$$r_{XY} = \frac{X_1^*Y_1^* + X_2^*Y_2^* + \dots + X_n^*Y_n^*}{n} \quad (29)$$

On "n" és la grandària de la mostra. Es pot comprovar que la quantitat determinada amb l'equació anterior es tracta d'un valor positiu si "X" i "Y" varien cap al mateix costat respecte les seves mitjanes en el mateix període, és a dir, mateix signe en un mateix període. I negatiu si la tendència es cap a costats diferents respecte les seves mitjanes en un mateix període, és a dir, diferent signe en un mateix període.

Generalment el valor del coeficient de correlació estarà entre -1 i 1, essent el primer el valor per relació lineal perfectament negativa, mentre que el segon indica una relació lineal perfectament positiva. En cas de que el resultat fos zero, no significaria que no hi existeix relació, sinó que simplement la relació de la mostra històrica que estem analitzant no és lineal.

Seguint amb els raonaments teòrics, la línia de regressió per predir "Y*" a partir de "X*" és una recta que òptimament passaria per l'origen si ambdós valors estandarditzats fossin de forma ideal normalment distribuïts, ja que té una pendent igual a la correlació entre "Y" i "X" que serà inferior a 1 en la gran majoria des casos. Amb tot, l'equació de la regressió en unitats estandarditzades és la següent:

$$\hat{Y}^* = r_{XY}X^* \quad (30)$$

Amb aquesta certa inclinació que s'obté del pendent de r_{XY} el que es pretén es reduir l'error quadràtic, ja que aquest es pren a partir de la diferència vertical entre les dades reals i les estimades. De fet, el terme regressió en si mateix es refereix a aquest aspecte, a la minimització de la suma dels errors quadràtics predint una variable respecte l'altra.

3.4.4.1. Model de regressió simple

En el cas de la regressió simple, el model de regressió comptarà amb una variable independent i amb dos paràmetres a ser estimats. És el model més simple de regressió i la seva equació és:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (31)$$

On les betes i l'èpsilon tenen les mateixes característiques descrites per al model general, com s'apria en l'equació (25) de l'apartat 3.4.4.

l'equació per a la predicció de "Y_t" a partir de la "X" queda doncs de la següent forma:

$$\hat{Y}_t = b_0 + b_1 X_t \quad (32)$$

On veiem que la constant estimada "b₀" es correspon al punt de tall amb l'eix y de la recta de regressió, és a dir quan x=0. També podem veure que el coeficient "b₁" és la pendent de la mateixa, que determinarà el canvi predit en "Y" per unitat de canvi en "X".

Pel mètode dels mínims quadrats podrem obtenir les expressions de càlcul dels dos coeficients "b", en que obtenim que el coeficient b₁ és igual a la correlació entre "X" i "Y" multiplicada per la relació de la desviació estàndard de "Y" respecte la desviació estàndard de "X". Matemàticament ho podem escriure també de forma conjunta com:

$$b_1 = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} \quad (33)$$

$$b_0 = \bar{Y} - b_1 \bar{X} \quad (34)$$

Una vegada tenim el model i els coeficients calculats, també podríem calcular els errors. A l'error estàndard del model s'hi sol referir com l'error estàndard de la regressió, que el podem calcular a partir de les fórmules vistes amb anterioritat, tenint en compte que es tindran dos graus de llibertat.

De l'error estàndard de la regressió també podríem calcular l'error estàndard en el càlcul dels paràmetres, que tot i que no hi farem referència directa, en podem dir que són directament proporcionals a l'error estàndard de la regressió i inversament proporcional a l'arrel de la grandària de la mostra, per tant, com més gran sigui la nostra mostra més gran serà la reducció de l'error en els nostres càlculs i la nostra regressió serà més precisa.

De forma recurrent és fa referència quan parlem de regressions al terme *R-squared*, aquest fa referència a la fracció de la variància de "Y" que es pot explicar a través del model de regressió simple. És a dir, és el percentatge pel qual la variància de la mostra dels errors (també anomenats residuals) és menor

que la variància de la mostra de "Y" en si mateixa, és igual al quadrat de la correlació entre elles. Aquest quadrat es correspon amb la *R-squared*:

$$R - squared = r_{XY}^2 \quad (35)$$

No obstant, en les futures determinacions s'haurà de tenir en compte tot l'esmentat de forma teòrica anteriorment i no caure en el parany de basar les nostres prediccions sola i simplement d'acord al valor de *R-squared* obtingut. A més, en la comparativa de models que farem en termes de bondat de l'ajust i dels errors, els errors que compararem seran altres ja esmentats amb anterioritat per tal que aquests siguin comuns a tots, agrupant-los de forma visual en taules, com veurem en apartats posteriors d'aquesta memòria.

3.4.5. Models ARIMA

Fins al moment s'ha vist diferents models de predicció de sèries temporals a partir de les pròpies dades històriques. Aquests models són aparentment independents amb les seves característiques i particularitats pròpies en el procediment estadístic, i per escollir entre uns i altres ens basarem en els errors que ens dona el model respecte les nostres dades, sempre assumint que les diferents hipòtesis que ens hem plantejat són correctes.

No obstant, tots aquests models podem dir que són casos especials d'uns models més generals de sèries temporals coneguts com *ARIMA models* o *Models ARIMA*. Aquest model neix com una adaptació a uns models de filtració de dades discretes temporals que van ser desenvolupats als anys trenta i quaranta per enginyers elèctrics, i que posteriorment van adaptar i desenvolupar els estadístics George Box i Gwilym Jenkins pel tractament de dades econòmiques als anys setanta, d'aquí que aquests models també siguin anomenats com *Box-Jenkins models*.

Les sigles del model ARIMA signifiquen *Auto-Regressive Integrated Moving Average*. Per entendre cadascuna de les components del nom del model ens serviran els diferents conceptes teòrics introduïts anteriorment dins de l'explicació dels models:

- Una sèrie temporal que necessita ser diferenciada per tal de convertir-la a estacionària la podem anomenar una sèrie integrada, d'aquí la "I" d'*integrated*.
- Els *lags* o retards de les sèries estacionaritzades són anomenats termes auto-regressius, d'aquí ve "AR" d'*auto-regressive*.
- Els *lags* o retards en els errors de predicció són anomenats els termes de mitjana mòbil, d'aquí la "MA" de *moving average*.

La particularitat principal d'aquests models és que reuneixen les característiques que fan bons els altres models, podent-los readaptar i modificar per obtenir més bondat en l'ajustament.

D'aquesta manera podem tenir els ajustaments fins que ens eliminen tota l'autocorrelació residual proporcionats pels models de camí aleatori generalitzats, podem incorporar les tendències a llarg termini o les periodicitats estacionals que ens permeten els models d'ajustament exponencial i també les propietats dels models de regressió estacionaritzats que utilitzen els *lags* o retards de les variables dependents i/o *lags* en els errors de predicció com a regressors. Els models més generals de predicció per sèries temporals poden estacionaritzats (totes les propietats estadístiques són constants amb el temps) per transformacions com la logarítmica o per diferenciacions.

Per diferenciar les diverses variants del model que estem utilitzant s'utilitza la terminologia ARIMA(p, d, q), on les lletres que acompanyen signifiquen el següent:

- p = nombre de termes auto-regressius del model.
- d = nombre de diferenciacions sense periodicitats estacionals.
- q = nombre de termes de mitjana mòbil.

Això ens donarà lloc al model ARIMA(p, d, q) utilitzat. Així mateix i tal com havíem vist en models anteriors, aquest tipus de models poden incloure una constant.

La determinació dels tres nombres anteriors serà la clau doncs per a l'elecció del model que millor s'ajusti a les dades que estem tractant. No obstant, també es el punt més difícil, és per això que seguir un bon procediment en la construcció del model té una gran importància. Els passos que de forma general podem assumir com els més adients per la construcció del model són els següents:

1. Determinar si la nostra sèrie temporal original necessita d'alguna transformació no lineal per tal de ser convertida o modificada a una forma on les seves variacions aleatòries locals siguin consistents en el temps i generalment simètriques pel que fa a l'aparença.
2. Notarem com "Y" la sèrie temporal que s'obté després del pas 1. Si aquesta "Y" encara no és estacionària, és a dir, existeix una tendència lineal, una tendència no lineal que varia de forma aleatòria, o sembla tenir una aparença de camí aleatori, llavors serà necessari aplicar una transformació consistent en diferenciar un primer cop les nostres dades. Això és, construir una nova variable que consisteixi en els canvis entre períodes de la nostra "Y".

3. Si l'aspecte és encara no-estacionari després de la transformació aplicada amb la primera diferenciació, s'haurà d'aplicar una transformació per diferenciació a la nova variable. És a dir, diferenciar un altre cop. Aquesta diferenciació serà inclosa en el terme "d" que hem definit anteriorment per anomenar el model. Generalment els seus valors seran 0, 1 o 2. Matemàticament les diferenciacions tindran la forma següent:

$$d = 0: \quad y_t = Y_t \quad (36)$$

$$d = 1: \quad y_t = Y_t - Y_{t-1} \quad (37)$$

$$d = 2: \quad y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} \quad (38)$$

4. Notarem "y" a la sèrie temporal estacionaritzada de la que hauríem de disposar arribats a aquest punt. En cas de que $d=0$, aquesta "y" coincidirà amb "Y". Una sèrie temporal estacionaritzada no té tendència, i la seva variància és constant en el temps, és a dir, que les seves autocorrelacions també seran constants en el temps. Llavors l'equació ARIMA per la predicció de "y" haurà de tenir una forma com la següent:

Predicció de "y" al període "t" = constant
 + suma ponderada dels últims "p" valors de "y"
 + suma ponderada dels últims "q" errors de predicció

Els valors de pes per la ponderació poden ser tant positius com negatius, i "p" i "q" es corresponen amb els termes definits anteriorment.

Si ho passem a llenguatge matemàtic l'expressió tindria l'aspecte següent:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (39)$$

On μ es correspon amb el terme de la constant, ϕ_k es correspon al coeficient del terme autoregressiu al lag "k", θ_k es correspon al coeficient de mitjana mòbil al lag "k" i $e_{t-k} = y_{t-k} - \hat{y}_{t-k}$ és l'error en la predicció fet al període t-k.

El fet que els termes autoregressius vagin acompanyats pel signe positiu i els termes de mitjana mòbil acompanyats pel negatiu és purament per conveni.

5. La predicció de la sèrie original al període "t", basada en les dades observades fins al període t-1, és obtinguda de la predicció de "y" desfent les vàries transformacions que se li han anat aplicant al llarg del procés, per tant, el últim pas és simplement desfer les transformacions. El procediment matemàtic de la reversió és el següent:

$$\text{Si teníem } d = 0: \quad \hat{Y}_t = \hat{y}_t \quad (40)$$

$$\text{Si teníem } d = 1: \quad \hat{Y}_t = \hat{y}_t + Y_{t-1} \quad (41)$$

$$\text{Si teníem } d = 2: \quad \hat{Y}_t = \hat{y}_t + 2Y_{t-1} - Y_{t-2} \quad (42)$$

D'aquesta manera doncs obtenim la predicció amb el model ARIMA, però en aquesta construcció anterior ja s'assumeix que coneixem el valor dels termes auto-regressius i de mitjana mòbil, que com hem dit planteja les major de les dificultats en l'elecció de model. Per tant, anem a parlar una mica més d'aquests dos termes.

El primer que en podem dir és que tot i que es plantegen de manera conjunta en l'explicació del model, no tenen perquè utilitzar-se els dos junts en la determinació del model. De fet, generalment no serà així i amb un dels dos termes ja serà suficient. Algunes sèries s'ajusten millor per termes autoregressius i d'altres per termes de mitjana mòbil, partint d'un cert nivell de diferenciació donat.

Una primera regla que podem aplicar de manera funcional per saber com és la nostra sèrie i quin terme hi tendirà a funcionar millor és la següent: si la sèrie estacionaritzada té una autocorrelació positiva al *lag* 1, en general els termes autoregressius funcionaran millor, mentre que si té una autocorrelació negativa al *lag* 1, els termes de mitjana mòbil en principi funcionaran millor.

Quan la sèrie tingui un comportament autoregressiu AR(1), el coeficient determina la rapidesa amb que la sèrie tendeix a tornar cap al seu valor mig. Quan aquest és un valor proper a zero, significa que hi torna de manera ràpida, mentre que si és un valor proper a 1, el retorn al valor mig es produirà més lentament. En el cas que hi hagi més coeficients, serà la suma dels diferents coeficients el que determinarà la velocitat de retorn al valor mig, i podria ser que la sèrie mostrés un patró oscil·latori.

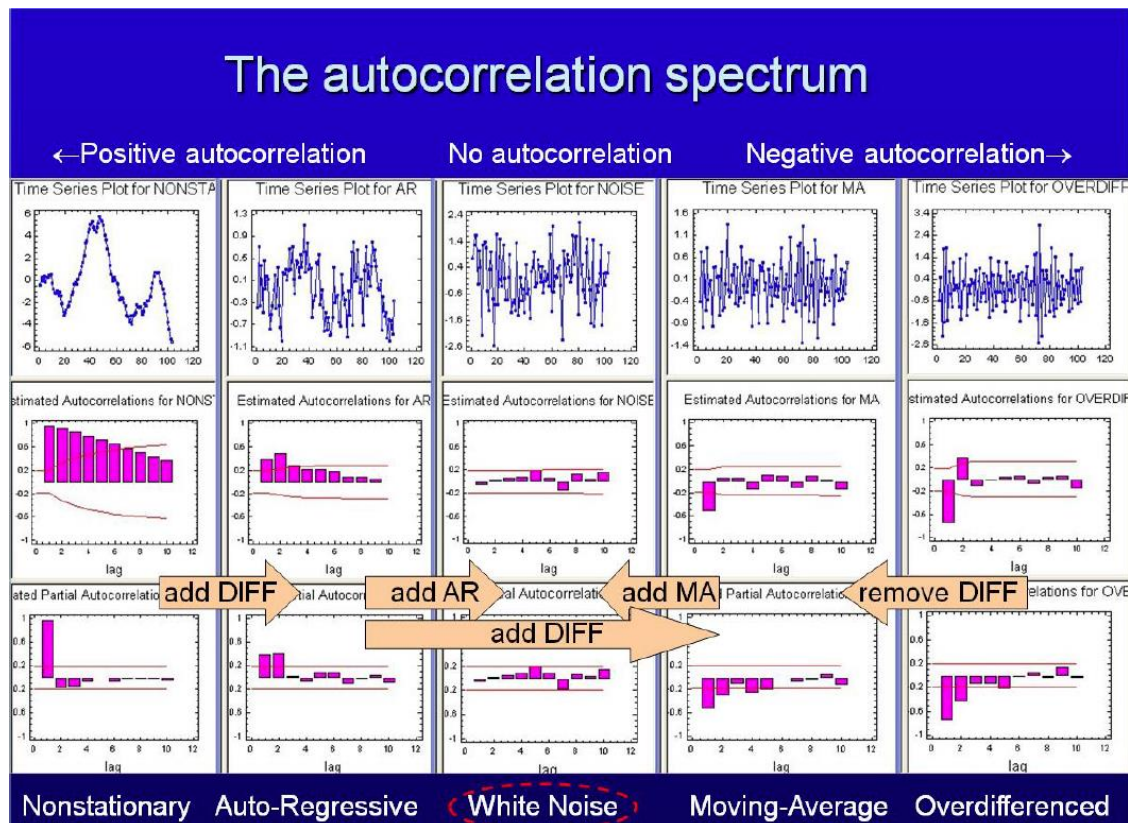
En canvi, quan una sèrie presenti un comportament de mitjana mòbil $MA(1)$, el coeficient indicarà la fracció de l'últim període que encara afecta al període actual. Si hi ha coeficient $MA(2)$, aquest indicarà la fracció dels dos períodes anteriors que encara afecta al període actual, i així successivament. En aquest cas hem de tenir en compte que els termes de mitjana mòbil es troben amb el signe negatiu, com hem vist en les fórmules anteriors.

Les eines que utilitzarem per identificar i analitzar els models ARIMA són bàsicament dos: els gràfics d'autocorrelació o *autocorrelation function* (ACF), i els gràfics d'autocorrelació parcial o *partial autocorrelation function* (PACF). De forma simplificada podem dir que el primer fa referència a l'autocorrelació de les sèries amb si mateixes en els diferents *lags*, i el segon mostra la quantitat de correlació al *lag* "k" que no està explicada per les autocorrelacions de baix ordre.

Observant els dos gràfics ens podem fer a la idea de quina es la tendència que marca la sèrie temporal que estem analitzant. En el cas de que la tendència sigui autoregressiva, veurem que el gràfic ACF es va reduint progressivament i que el PACF presenta un canvi brusc al cap de pocs *lags*. A més, generalment presentaran autocorrelació positiva al primer *lag*. En canvi, en el cas que la tendència sigui de mitjana mòbil, podrem veure que el gràfic ACF té un canvi brusc al cap de pocs *lags*, mentre que el PACF s'anirà reduint progressivament, a més en aquest cas generalment presentaran autocorrelació negativa al primer *lag*. Per tant, veiem que pràcticament una tendència de comportament és contrària de l'altra.

En el procés de determinació de la tendència de la sèrie serà important la diferenciació de la mateixa, ja que el comportament pot variar segons quin sigui el seu grau de diferenciació. Com a criteri general, les sèries sense diferenciació presenten autocorrelació positiva, i, per tant, tendència autoregressiva, mentre que després de una o més diferenciacions l'autocorrelació comença a ser més negativa, i la tendència de mitjana mòbil es va fent més present. De tota manera, com sempre buscarem el model més simple, s'ha de vigilar no sobre diferenciar quan una sèrie ja presenta autocorrelació zero o negativa al primer *lag*.

El gràfic següent pot ser de gran utilitat a l'hora de prendre decisions per ajustar el model:



Il·lustració 5 Pautes d'ajust del model ARIMA en funció dels residuals i les autocorrelacions (Nau, 2014)

En relació amb tot l'esmentat anteriorment, havent explicat la construcció del model, les bases en les que es fonamenta, i certes pautes d'utilització, sol ens queda doncs afegir els passos a seguir per fer el correcte ajustament del mateix:

1. Determinar l'ordre de diferenciació.
2. Determinar els nombres dels termes auto-regressius i de mitjana mòbil "AR" i "MA" (que es corresponen amb els valor de "p" i "q").
3. Ajustar el model de manera que ens quedi correcte en els diferents gràfics i paràmetres, obtenint el que s'anomena el *white noise*, tal i com es pot veure en la Il·lustració 2. En cas que no sigui així caldrà repetir els primers dos passos.

Amb tot, veiem que un dels punts claus del quart pas per la construcció del model serà fixar-se en els gràfics ACF i PACF de "y" i veure quina és la tendència clara de la nostra sèrie.

Una vegada identificat el model, haurem de fixar-nos que el terme "AR" o "MA" d'ordre major sigui significativament diferent de zero d'acord als estàndards típics dels models regressiu, és a dir, que el *t-stat* sigui més gran de 2 en magnitud i que el *p-value* sigui menor de 0.05. Per realitzar aquesta comprovació sol cal mirar el terme d'ordre major, ja que els d'ordre petit ja s'hi inclouran.

Tenint clares aquestes pautes representades en la figura anterior es pot trobar el model més simple que funcioni amb la nostra sèrie. Entenent per simples els models en que la suma dels termes auto-regressius i de mitjana mòbil (AR i MA) sigui menor que 3, és a dir, que “p” i “q” sumin menys de 3. Generalment aquesta apreciació serà suficient, ja que incrementar els termes en valors més alts afegeix complexitat al model que moltes vegades no és necessària.

Alguns dels models més comuns que poden ser explicats per models ARIMA i ja tractats en aquesta memòria són per exemple:

$ARIMA(0,0,0)+c$ = model de la mitjana

$ARIMA(0,1,0)$ = model del camí aleatori

$ARIMA(0,1,0)+c$ = model del camí aleatori amb constant de tendència (model del camí aleatori geomètric si apliquem una transformació logarítmica)

$ARIMA(0,1,1)$ = model d'ajustament exponencial simple

$ARIMA(0,1,1)+c$ = model d'ajustament exponencial simple amb constant de tendència

Com veurem més endavant en la memòria, algun d'aquests models i d'altres que també podem crear amb ARIMA seran escollits per ajustar les nostres sèries temporals i fer les comprovacions adients. Això és degut a que com ja s'ha explicat en aquest apartat és un model prou complex com per aglutinar les diferents característiques dels models anteriorment explicats, i amb el domini d'aquests ARIMA podem observar patrons útils des d'una sola perspectiva que després ens pot portar a conclusions més globals respecte l'anàlisi de les sèries temporals que tractarem.

3.5.PERÍODES DE VERIFICACIÓ

Estrictament parlant tots els models seran analitzats en funció del bé que s'han ajustat al passat, no en funció del bé que hauran predit el futur. És cert que en principi si volem tenir una bona predicció ens hem de fixar en com els models han anat responent al passat, ja que en principi si res canviés el model prediria de forma prou representativa com per considerar que és la forma correcta que tindrà el futur.

No obstant, s'haurà de fer més proves per obtenir finalment quin serà el millor model a utilitzar. Una de les formes més simples però eficaces serà treure una part de les nostres dades per tal de tenir un període de verificació del nostre model en relació a les dades que prediria del futur, les quals ja coneixem.

D'aquesta manera no solament podrem veure si el model ha funcionat, sinó també com han canviat els paràmetres estimats, ja que potser quan hem tret part de les dades el millor model en termes d'errors és un altre al que teníem anteriorment.

En el cas que el model surti similar al que ha esdevingut els valors reals, significarà que les hipòtesis prèvies per l'elecció del model han estat bones, i també que els valors específics dels paràmetres que han estat estimats de la mostra també són bons.

En la comparació dels errors, a vegades pot passar que el RMSE surti més gran en el període de verificació, això pot ser indicador que les dades necessiten d'alguna transformació prèvia, tenint en compte que per calcular posteriorment un altre cop l'error, aquesta transformació cal desfer-se. En els períodes de verificació generalment és millor comparar els valors del MAPE per veure si el model ha funcionat bé en la fase de verificació i en la d'estimació, ja que aquest sol ser igualment comparable. Altrament, i com s'ha comentat amb anterioritat, també s'observarà el valor del MAE per obtenir una visió de conjunt més gran.

En el present estudi també es tindran en compte aquests canvis i es valoraran els models en funció de dos períodes de verificació, un amb un termini més curt respecte la realitat, i un amb un període a més llarg termini respecte la realitat. Aquests períodes ens seran molt útils a l'hora de treure conclusions i discutir els diferents resultats obtinguts.

4. RESULTATS

Totes les dades obtingudes per a la realització del treball han estat obtingudes dels diferents estaments i institucions estatals que les ofereixen de forma fiable i contrastada. Això és, que han estat obtingudes directament de l'estament oficial *Puertos del Estado*, tant dels seus anuaris estadístics com dels registres d'estadístiques de trànsit que s'hi ofereix (Puertos.es, n.d.).

Aquestes dades s'han complementat amb les dades aportades també de forma oficial per les Autoritats Portuàries dels ports analitzats i amb els diversos BOE que oferien informació respecte el tràfic portuari (Portdebarcelona.cat, Porttarragona.cat, Valenciaport.com, n.d.)

S'ha de tenir en compte que les dades que utilitzarem són fiables i estan registrades des de l'any 1973 fins el 2015, les quals segmentarem com veurem posteriorment en dos períodes de verificació.

En tots els models la unitat que farem servir per la mesura i la predicció de tràfic portuari seran els TEU, acrònim del terme anglès *Twenty-foot equivalent unit*, que representa la unitat de mesura de capacitat inexacta del transport marítim expressada en contenidors.

Els ports que s'analitzaran amb els mètodes anteriorment explicats seran el de Barcelona, el de Tarragona i el de València, als que ens hi referirem com BTV.

Tot i que inicialment l'estudi estava pensat per l'anàlisi singular del cas de Barcelona, s'ha decidit ampliar aquest estudi als ports de Tarragona i València pel fet que també s'inclouen dins de les 10 Autoritats Portuàries amb més import de xifra de negocis ICN de l'Estat, i és interessant l'anàlisi de cara a posar en valor tot el que esdevé aquesta zona superior del corredor mediterrani a nivell de transport cap a l'interior i l'exterior peninsular.

Tarragona ens aportarà conclusions molt interessants dels models que estem tractant per les seves dades històriques més singulars respecte els altres dos casos, a més de ser un cas també català. I València, que juntament amb *Bahía de Algeciras* i Barcelona conforma els més importants de l'Estat a nivell de volum de negoci, ens permet analitzar un altre dels ports més potents i també concentrar informació d'un altre cas proper als altres dos, d'on es treuran conclusions útils d'aquesta zona superior del corredor.

Aquests ports s'analitzaran tant de forma individual com finalment en conjunt tenint en compte tots els TEU que s'hi mouen entre els tres.

El procediment a seguir en aquest apartat serà el mateix per cada port. Primerament analitzarem les dades de forma global, observant les tendències durant les dades històriques.

A continuació, es passarà a fer una regressió simple del model, que ens servirà com a mètode més simple per a analitzar les nostres dades.

Posteriorment, en relació a aquest mètode simple es comprovaran els altres i diferents mètodes i combinacions exposats anteriorment dins del que és l'anàlisi de sèries temporals. Totes aquestes anàlisis més complexes es materialitzaran a través dels diferents ARIMAS, que com hem vist poden englobar els diferents models tractats a través de les seves múltiples combinacions. Amb tot, s'extrauran els valors dels errors amb cada model i combinació.

Dins d'aquesta anàlisi s'hi inclourà els tests de verificació dels models ja introduïts anteriorment, que consistiran en la selecció de diferents subconjunts de dades per veure com respon el model respecte dades que ja ens seran conegudes, això ens permetrà realment comprovar quin ha estat funcionant millor en les dades i tràfic actual.

Amb aquest procediment serà possible analitzar els errors dels models respecte la realitat des de dos perspectives temporals diferents: un període més proper a partir del 2010 i un més a mig-llarg termini des del 2005. La comparació de prediccions i de resultats respecte les dades que ja coneixem fins el 2015 ens donarà un registre dels errors ja esmentats i ens permetrà veure el funcionament dels models a curt i llarg termini, el qual ens portarà a conclusions interessants.

D'aquesta manera amb els diferents models tractats i veient quins són els que a priori funcionen o s'adapten més a la realitat, podrem veure quina és la combinació o mètode que d'acord a les característiques del port funcioni de forma òptima.

Degut a què les dades segueixen tendències i no són constants, es tindrà en compte el fet que el model disposarà de constant, sempre i quan el diagnòstic del model no la consideri significativa.

L'ordre dels models i els objectius previs en l'ús de cadascun en la nostra anàlisi seran el següents:

- Evolució temporal dels ports analitzats

En aquest primer apartat s'analitzarà globalment el tipus de dades i l'evolució temporal dels TEU en cada cas per tal de poder extreure primeres valoracions respecte quin podrà ser el funcionament dels mètodes utilitzats en cadascun dels casos.

- Regressió simple

El model de regressió lineal és el més simple de tots i ens servirà com a punt de partida per comparar el que seran els diferents models complexes d'ARIMA amb aquest. Es veurà si l'ús d'un model simple funciona de forma auxiliar per entendre les tendències generals de les dades, o si per el contrari té un ajust que sumat a la seva simplicitat fa que sigui un model a tenir en compte.

- Model de la mitjana (ARIMA(0,0,0)+c)

És el model més simple que es pot construir amb els models ARIMA, i més que esperar grans resultats del model, s'espera que serveixi com un primer test o presa de contacte amb aquests tipus de models, que al mateix temps servirà també com a lligam amb molts conceptes que tractats de forma teòrica amb anterioritat. A més, es podrà observar la tendència de les dades pel que fa als estadístics i als gràfics que ho indiquen.

- Model de camí aleatori amb constant (ARIMA(0,1,0)+c)

El model de camí aleatori o *random walk* en anglès es correspon a aplicar un primer ordre de diferenciació al model de la mitjana anterior. Aquesta transformació s'hauria de fer present sobretot en una millora en els gràfics de residuals, i ha de servir per seguir observant quina tendència predomina en la tipologia de les dades.

- Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)

En aquest model s'augmentarà en un grau el terme autoregressiu dels models ARIMA, on la característica principal serà l'intent d'incrementar l'ajust estadístic dels models observats fins al moment. S'hi comprovarà si aquest model de tendència autoregressiva funciona bé aplicat als diferents casos, essent el més simple dels models autoregressius que s'hi pot construir.

- Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)

Dins d'aquest model es provarà si es combinen els punts que cadascun dels dos models anteriors tenien a favor, d'una banda la transformació de les dades amb la diferenciació, i de l'altra l'ordre autoregressiu aportat prèviament.

- Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)

S'analitzarà aquest model per tal de comprovar si un terme autoregressiu més pot acabar de polir els models autoregressius anteriors. A més, es tracta d'un model que en els diferents àmbits en que s'analitzen sèries temporals sol funcionar correctament. Tot i això cal recordar que com més simple millor, i si la diferència respecte als altres no és molt gran, serà millor opció decantar-se per un dels anteriors.

- Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)

El model d'ajustament exponencial funciona bé en molts dels àmbits on s'analitzen sèries temporals, de forma que és interessant provar com funcionarà un model de mitjana mòbil amb les dades que disposem, per tal de comparar-lo amb els anteriors i ampliar el ventall d'opcions tractades.

Cadascun d'aquests es presentaran els respectius apartats següents, que es separaran en: Descripció de les dades, Estimacions i prediccions i Exercicis de verificació.

En el primer es mostrarà l'evolució temporal de cada port, esmentant ja possibles particularitats de les dades en els ports analitzats.

En el segon es presentaran les estimacions i prediccions de cadascun dels models, les quals en el cas dels models ARIMA aniran acompanyats dels gràfics d'autocorrelació (*auto correlation function*, ACF) i dels gràfics d'autocorrelació parcial (*partial autocorrelation function*, PACF), com també del gràfic de residuals de cada model. Aquests aportaran informació estadística de cada model que permetrà observar tant si un model és acceptable des del punt de vista teòric, com extreure argumentacions respecte possibles modificacions amb l'objectiu de millorar l'ajust que s'hi obté de cadascun.

Aquests gràfics es presentaran solament en el cas de l'apartat amb dades originals degut a que en els períodes de verificació seran idèntics per cada model però simplement amb menys dades en el cas dels residuals, i idèntics en els ACF i PACF però simplement amb els canvis presentats en un o dos *lags* més propers. Aquests no serien de gran importància ja que com s'ha vist de forma teòrica interessarà bàsicament el comportament dels primers *lags* o retards. Per tant, es presentaran sol els dels originals a fi d'economitzar espai i fer-ho més versàtil i entenedor.

Finalment, en el tercer apartat es presentaran les prediccions dels models en els dos períodes de verificació (2010 i 2005) i les taules dels errors amb valors arrodonits dels models anomenats anteriorment en cadascun dels ports analitzats, que es podran extreure comparant les prediccions amb les dades originals fins l'any 2015. Això permetrà fer valoracions respecte la precisió en les estimacions dels diferents models, primerament de forma gràfica, i, posteriorment, amb les taules dels errors que es presentaran per en l'apartat de discussions per tal d'aprofundir-hi juntament amb tota l'informació disponible.

4.1. DESCRIPCIÓ DE LES DADES

4.1.1. Barcelona

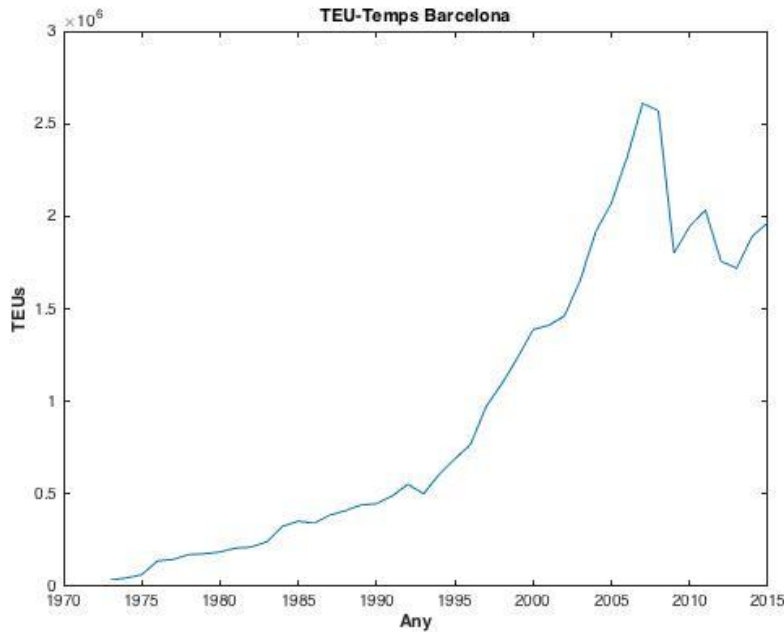


Figura 1 Evolució temporal del nombre de TEUS Barcelona

Es pot veure que en el cas de Barcelona les dades són lineals en el seu creixement fins arribar al període de crisi econòmica, on just abans a aquest període presenta el seu pendent màxim de creixement. La crisi es manifestarà en tots els casos ja que influeix de forma directa en el creixement dels TEU.

De fet s'hi observa una disminució clara que després s'estabilitza amb oscil·lacions al voltant dels 1.8M TEU. Els models presentaran a priori dificultat en la predicció d'aquest canvi de tendència, s'analitzarà posteriorment amb la presentació de més resultats.

4.1.2. Tarragona

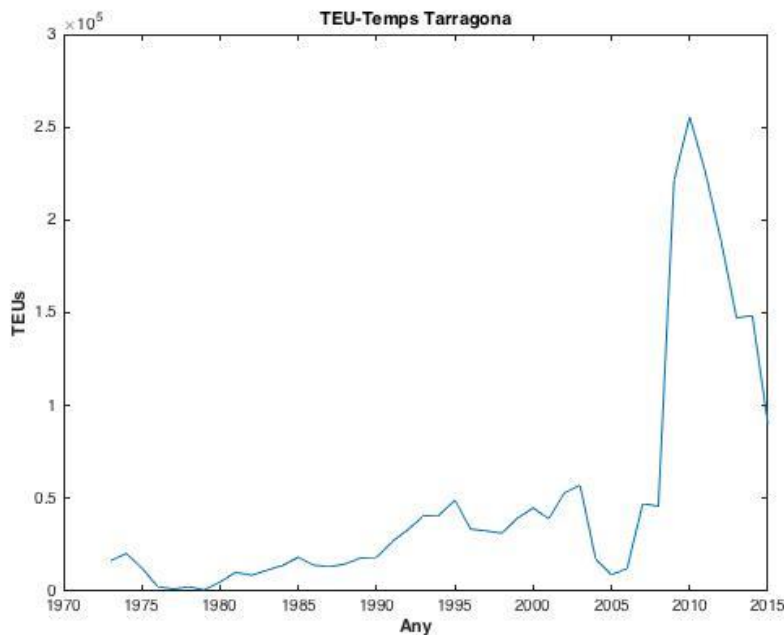


Figura 2 Evolució temporal del nombre de TEUS Tarragona

El cas de Tarragona és de llarg el que presenta una distribució de dades menys favorables pels models tractats en aquesta memòria.

S'hi observa clarament dos períodes pel que fa al nombre de TEU. Un primer fins l'any 2003 aproximadament, on la tendència és molt lineal i el creixement molt poc perceptible.

A partir d'aquest any hi ha la davallada màxima el 2005 i després comencen els canvis més grans que observarem. Hi trobem un creixement amb una pendent pràcticament vertical en els períodes previs a la crisi, quintuplicant els TEU de forma sense precedents. Així mateix també experimenta una davallada en el nombre de contenidors molt gran després de la crisi.

És sense dubte el cas més damnificat per aquesta crisi econòmica, tant per l'aspecte positiu abans de la crisi, com per la decaiguda de TEU posterior.

S'haurà de veure si els models univariants seran suficient per captar aquests dos forts canvis de tendència que s'hi presenten.

4.1.3. València

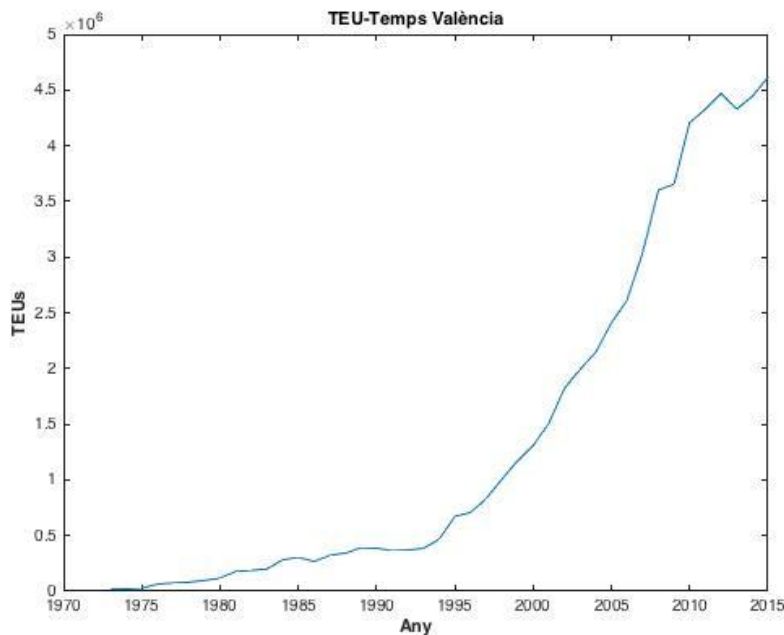


Figura 3 Evolució temporal del nombre de TEUS València

El cas de València és el més continu de tots els que s'analitzarà, a nivell que no s'hi observa cap canvi brusc pel que fa al nombre de TEU.

És una evolució que pot ser molt explicativa pel que fa al procés de containerització que descriuen Guerrero i Rodrigue (2014). S'hi aprecia una primera pendent suau, seguida d'una de major inclinació pel que fa al creixement que finalment acaba en una tercera pendent una mica més suau on s'hi torna a tenir un creixement més equilibrat.

En aquest cas la crisi es pot observar però de forma menys abrupta, ja que sol s'hi veu una petita davallada que sembla corregir-se en els anys posteriors.

Pel tipus de model que es tracta en aquest treball, València pot ser un cas on les precisions obtingudes siguin altes, ho comprovarem en els diferents períodes de verificació.

4.1.4. BTV

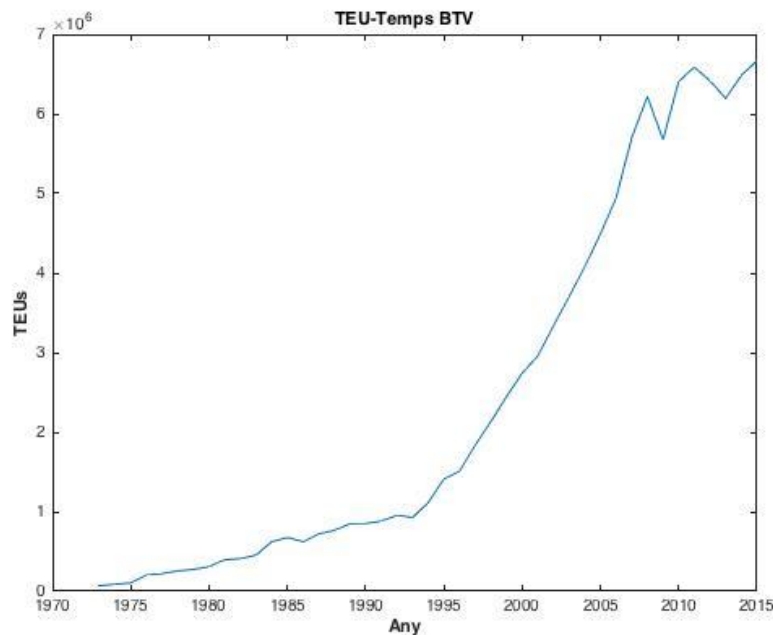


Figura 4 Evolució temporal del nombre de TEUS BTV

En el cas conjunt dels tres ports BTV (Barcelona, Tarragona i València) s'hi observa en conjunt el que s'ha pogut apreciar en els altres casos.

Primerament les dos primeres pendents clarament diferenciades que es poden correspondre a la pròpia evolució de la containerització en si mateixa.

Posteriorment si és cert que en global hi hauria una tercera pendent estabilitzadora, però també és cert que s'hi observa els daltabaixos provocats pel període de crisi en els diferents casos, que es veu representat en dos caigudes en el creixement durant aquest període.

Entendre bé les prediccions que surten d'aquest cas ens ajudarà a tenir una idea global de la possible evolució que pot tenir aquesta zona superior del corredor mediterrani a nivell peninsular.

4.2. ESTIMACIONS I PREDICCIONS

Les estimacions i prediccions mostrades en aquest apartat són respecte les dades originals sense extracció de dades en períodes de verificació. Així mateix, es presentaran els gràfics de residuals i d'autocorrelacions dels diferents models ARIMA, els quals seran comentats de forma singular per a cada model d'acord amb els objectius que s'han marcat en l'ús dels diversos models.

Aquests primers apunts teòrics sobre la bondat estadística i funcionament de cada model estaran lligats amb les verificacions posteriors i amb les argumentacions en els apartats tant de discussió com de conclusions.

4.2.1. Regressió simple

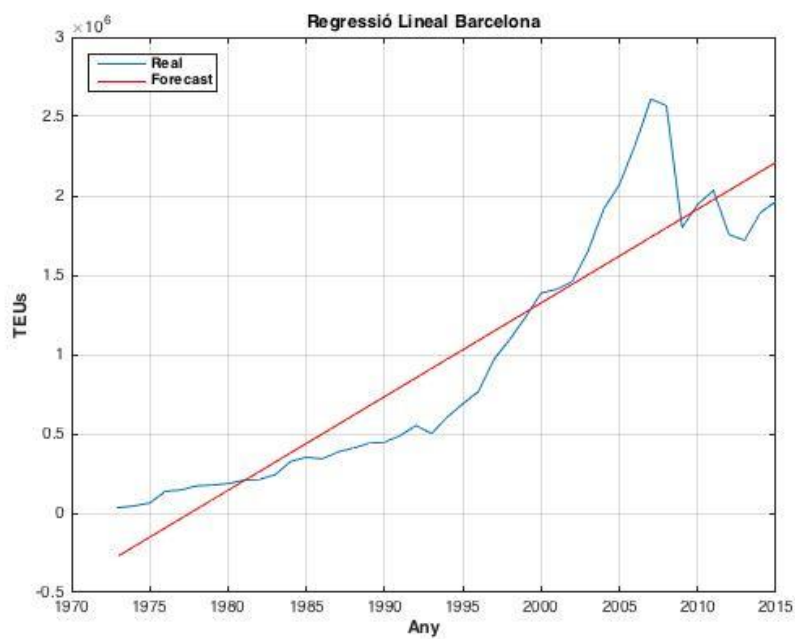


Figura 5 Regressió simple Barcelona

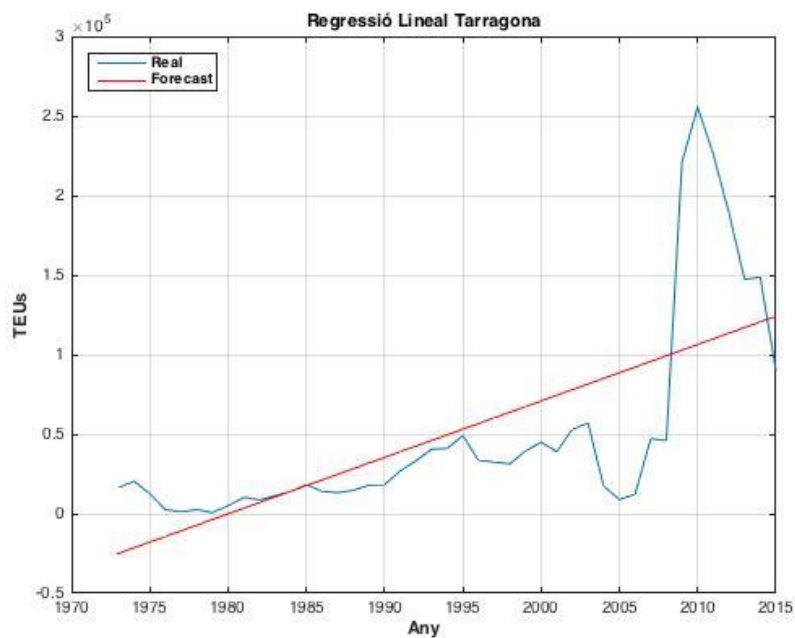


Figura 6 Regressió simple Tarragona

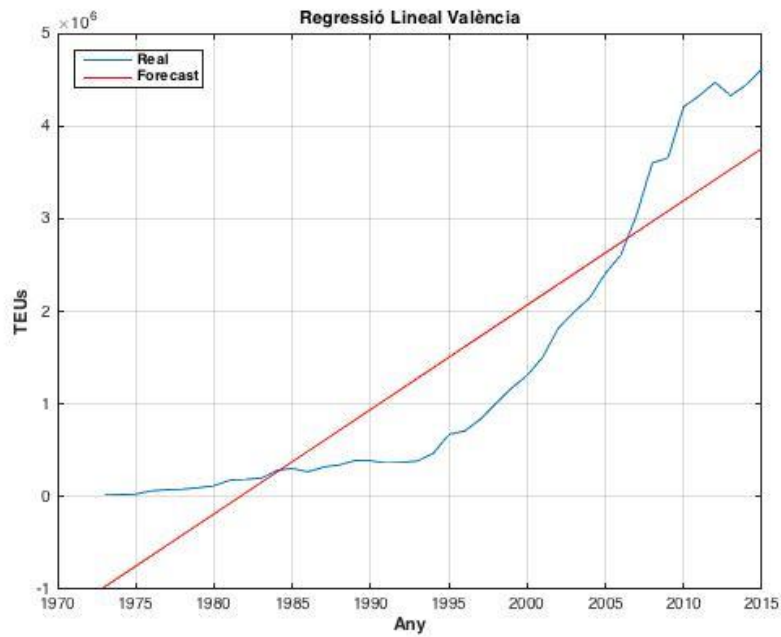


Figura 7 Regressió simple València

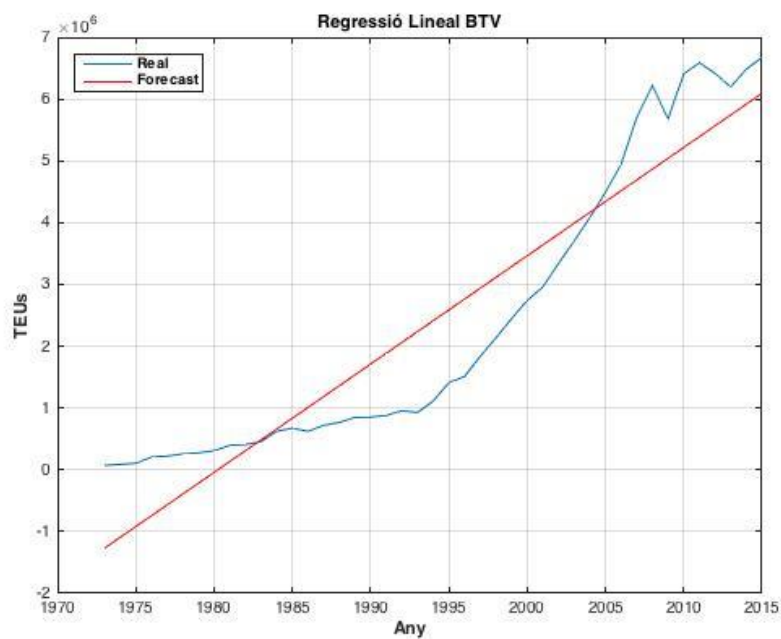


Figura 8 Regressió simple BTV

4.2.2. Model de la mitjana (ARIMA(0,0,0)+c)

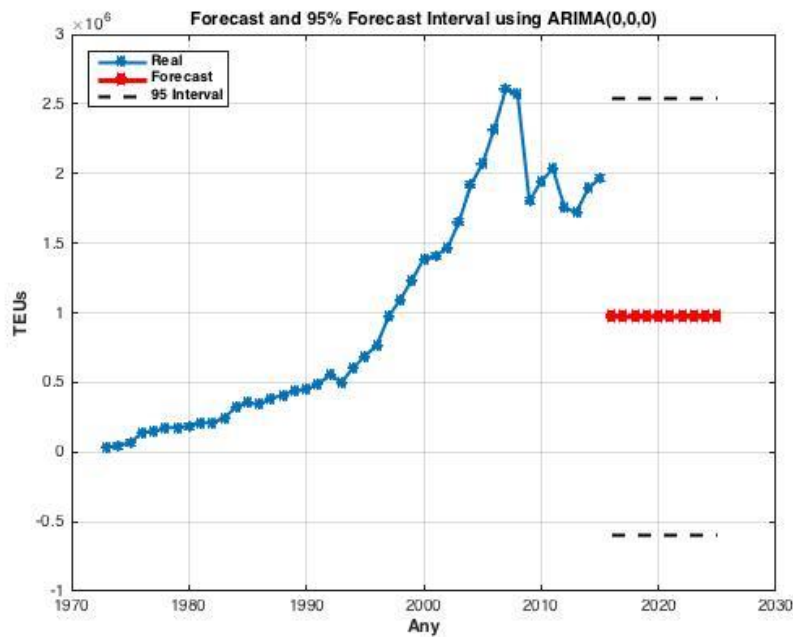


Figura 9 Model de la mitjana Barcelona

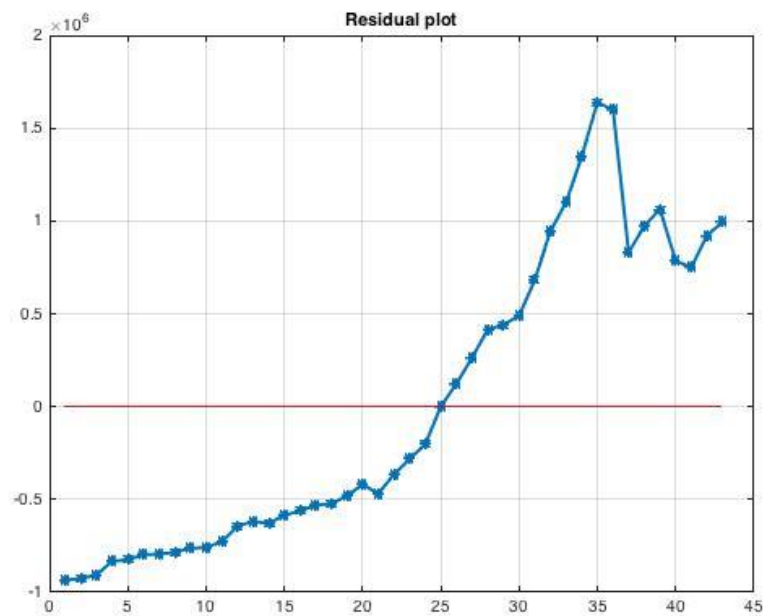


Figura 10 Model de la mitjana Barcelona gràfic de residualls

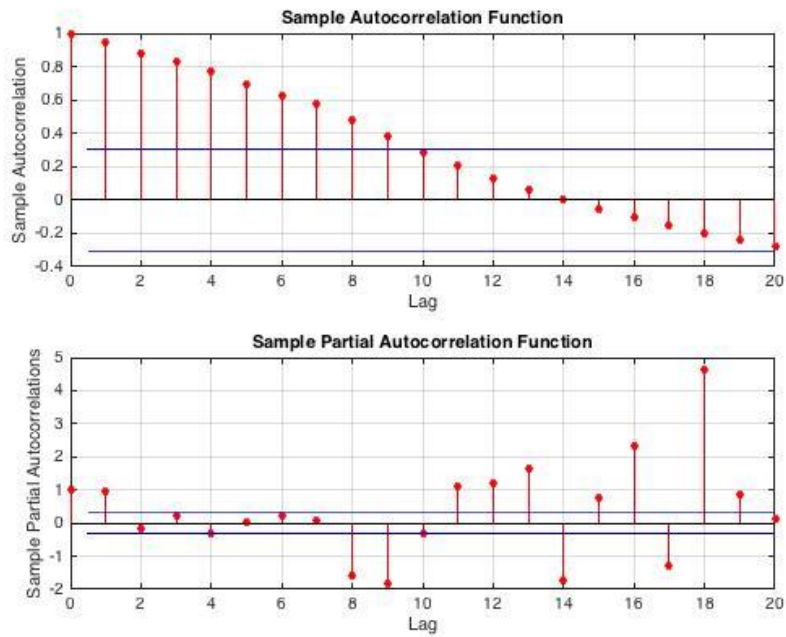


Figura 11 Model de la mitjana Barcelona gràfic d'autocorrelacions

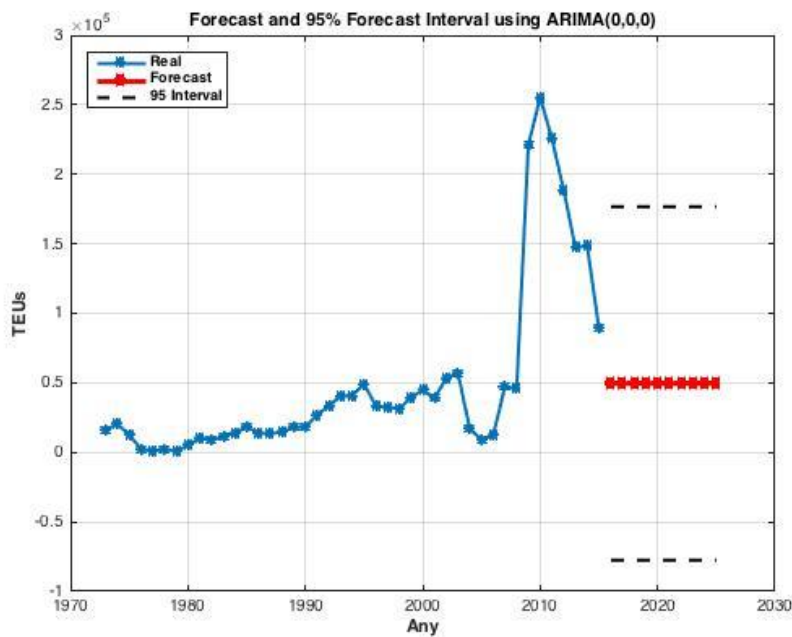


Figura 12 Model de la mitjana Tarragona

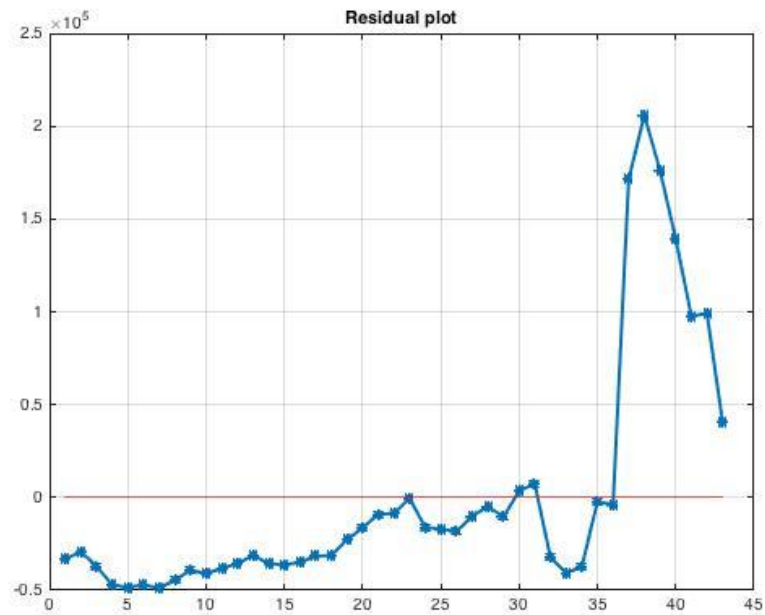


Figura 13 Model de la mitjana Tarragona gràfic de residualls

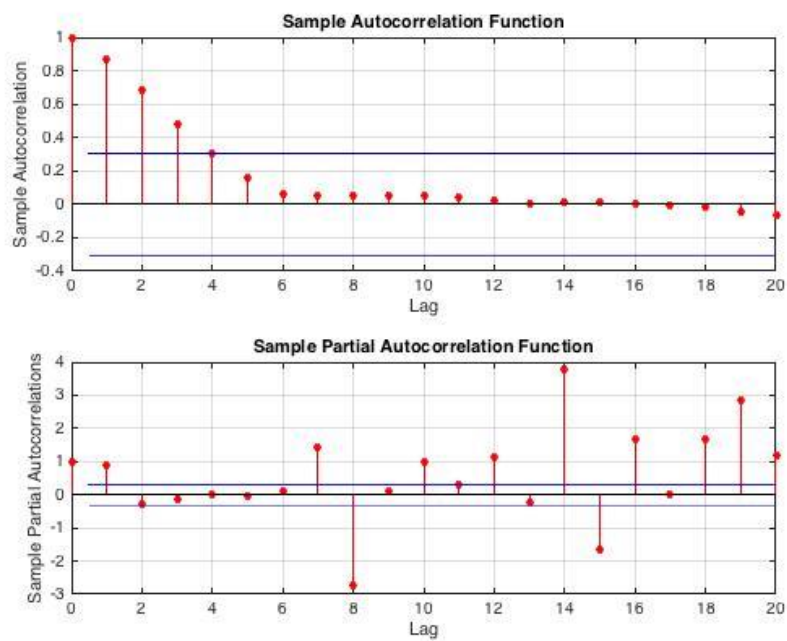


Figura 14 Model de la mitjana Tarragona gràfic d'autocorrelacions

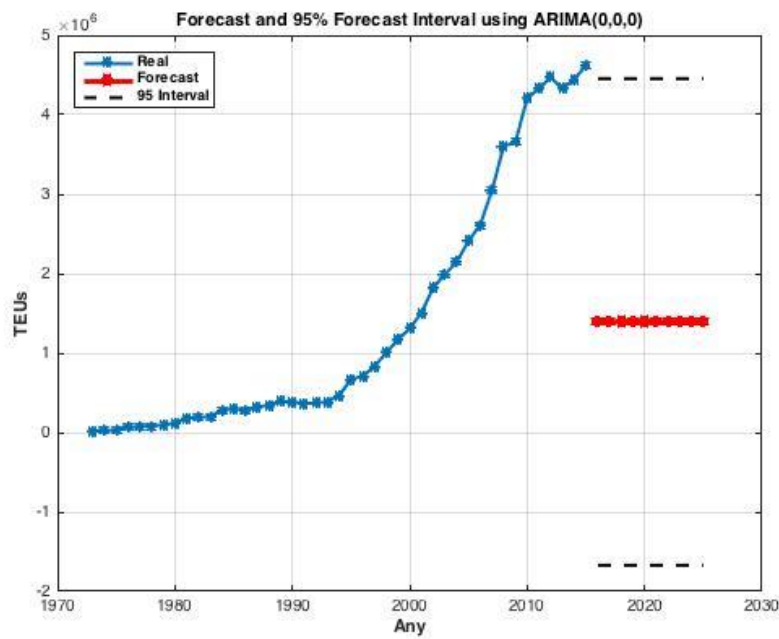


Figura 15 Model de la mitjana València

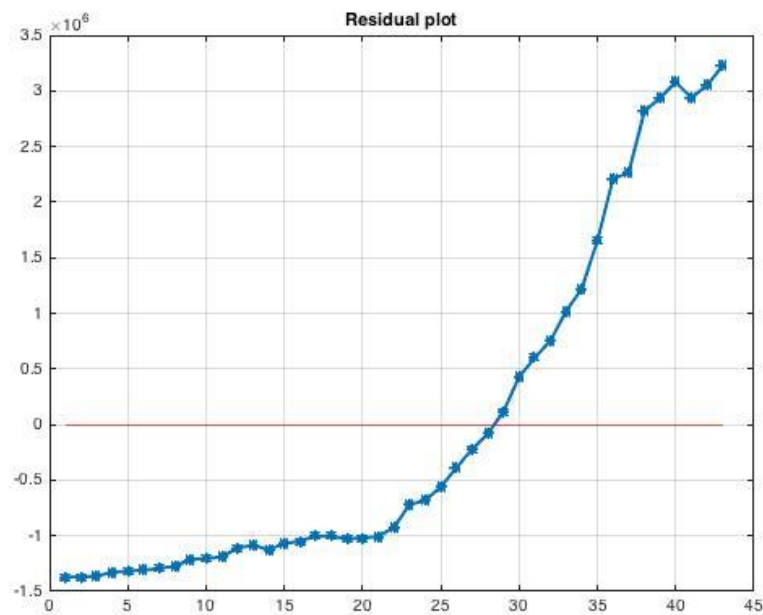


Figura 16 Model de la mitjana València gràfic de residuals

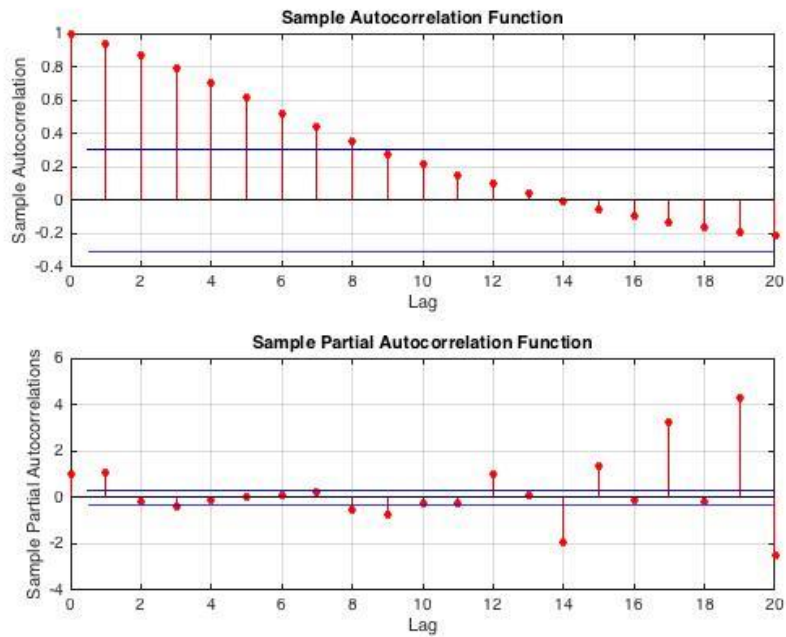


Figura 17 Model de la mitjana València gràfic d'autocorrelacions

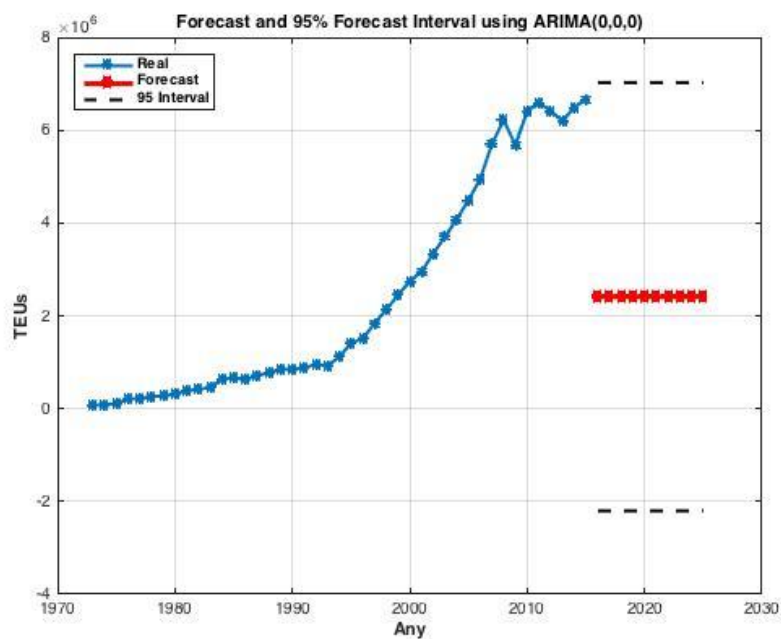


Figura 18 Model de la mitjana BTV

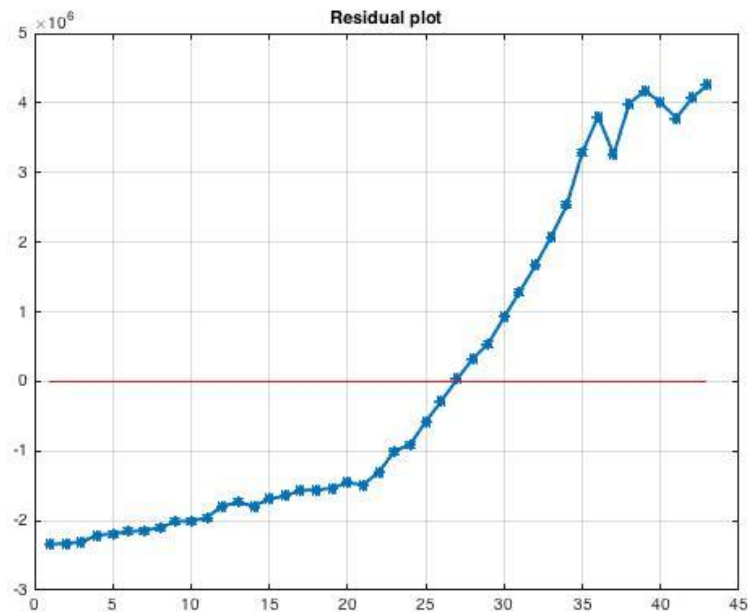


Figura 19 Model de la mitjana BTV gràfic de residuals

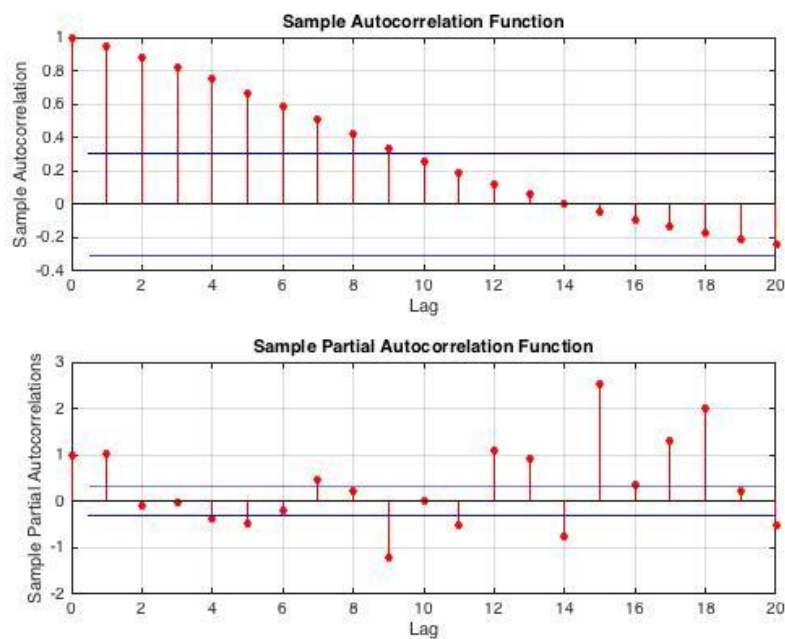


Figura 20 Model de la mitjana BTV gràfic d'autocorrelacions

En tots els casos s'hi segueix una sèrie de patrons que venen donats per la simplicitat del model en si.

Es pot observar que el gràfic de residuals no segueix un patró aleatori, per tant, seria necessària una diferenciació de les dades per tal de corregir aquesta mancança.

Pel que fa als gràfics d'autocorrelació, exceptuant el cas de Tarragona en què no està tan remarcant, presenten una tendència autoregressiva per la caiguda lenta dels ACF i presentar autocorrelacions positives, que com s'ha vist de forma teòrica en són indicadors clars d'aquesta tendència.

4.2.3. Model del camí aleatori (ARIMA(0,1,0)+c)

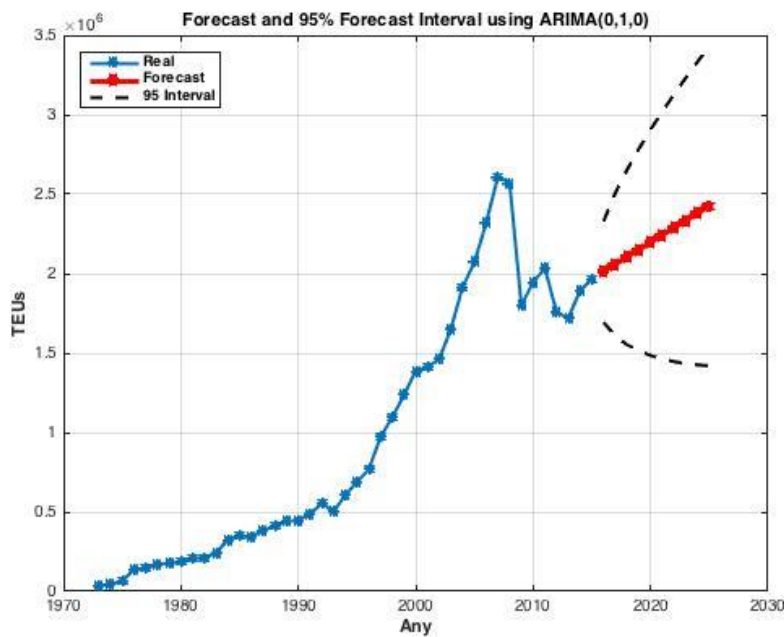


Figura 21 Model del camí aleatori Barcelona

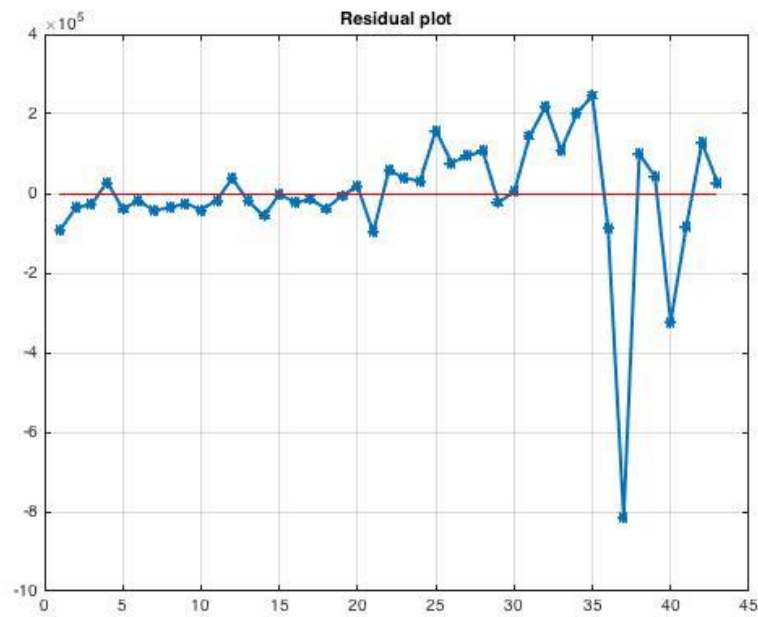


Figura 22 Model del camí aleatori Barcelona gràfic de residuals

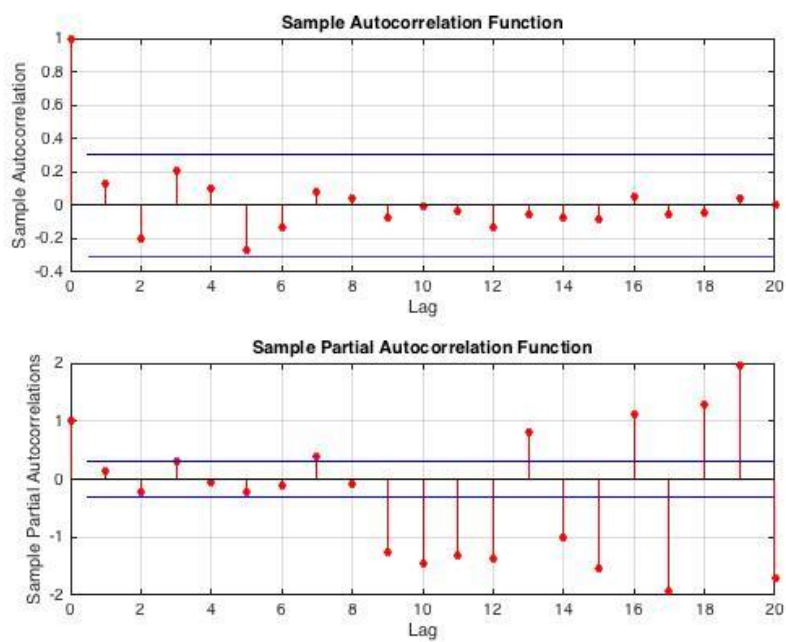


Figura 23 Model del camí aleatori Barcelona gràfic d'autocorrelacions

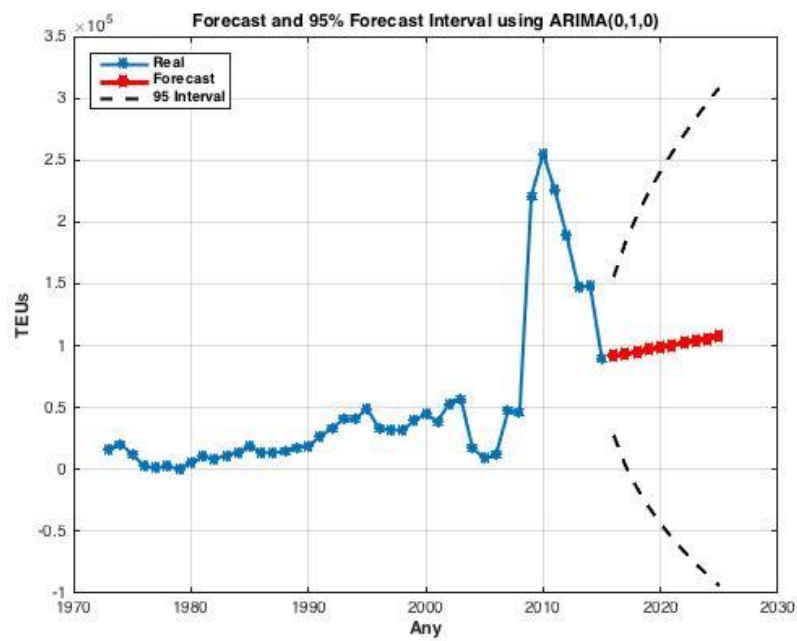


Figura 24 Model del camí aleatori Tarragona

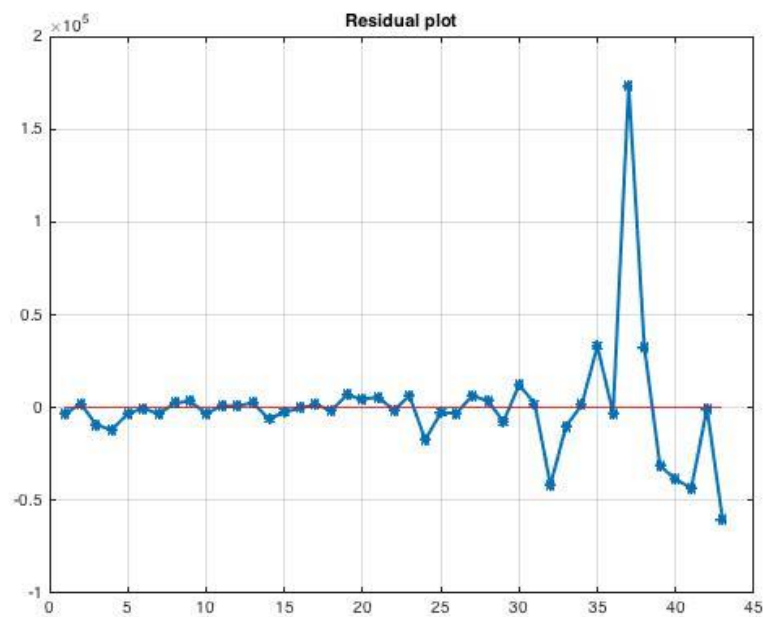


Figura 25 Model del camí aleatori Tarragona gràfic de residualls

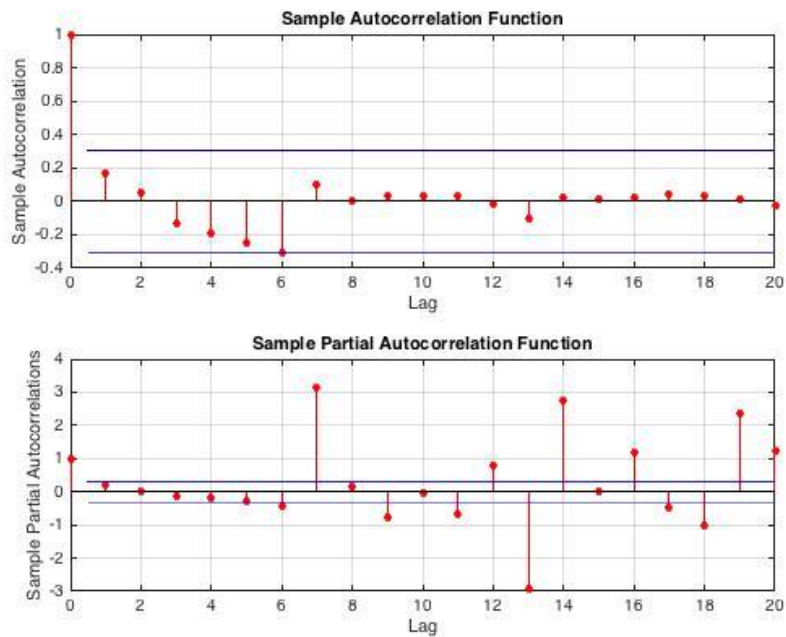


Figura 26 Model del camí aleatori Tarragona gràfic d'autocorrelacions

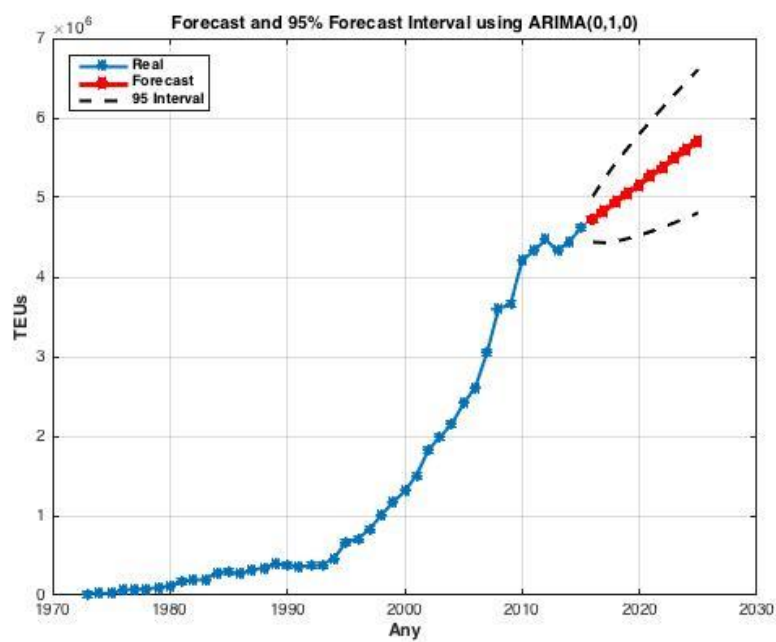


Figura 27 Model del camí aleatori València

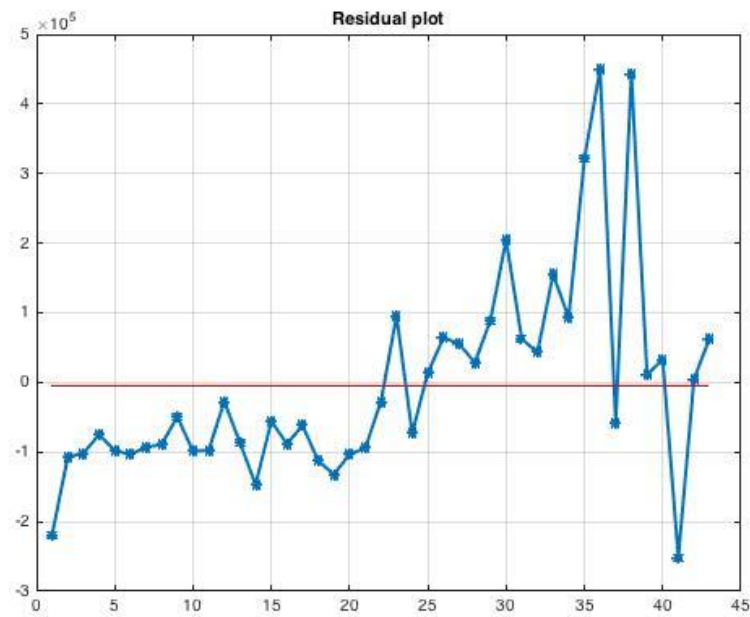


Figura 28 Model del camí aleatori València gràfic de residuals

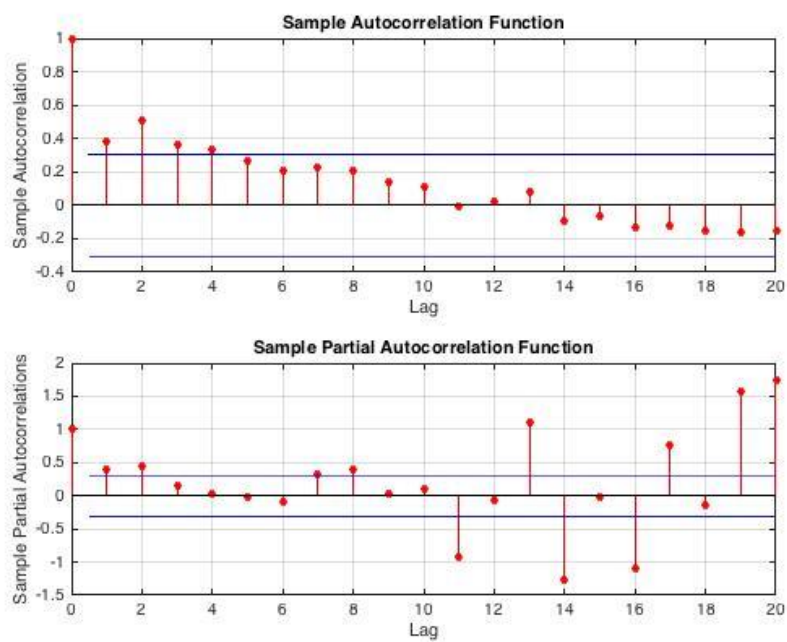


Figura 29 Model del camí aleatori València gràfic d'autocorrelacions

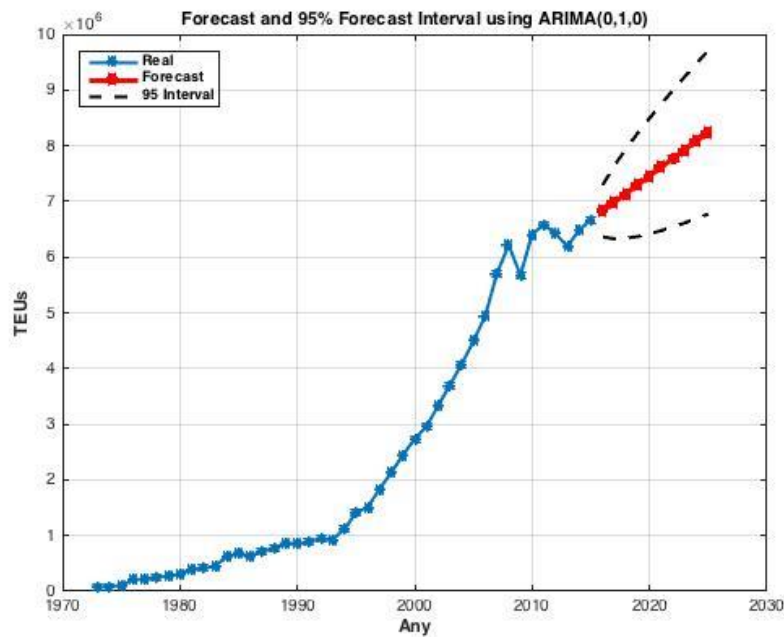


Figura 30 Model del camí aleatori BTV

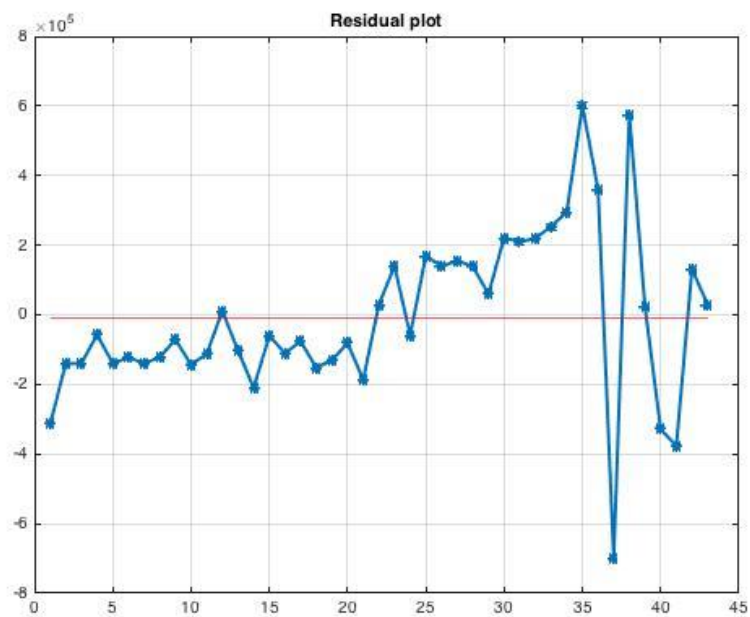


Figura 31 Model del camí aleatori BTV gràfic de residuals

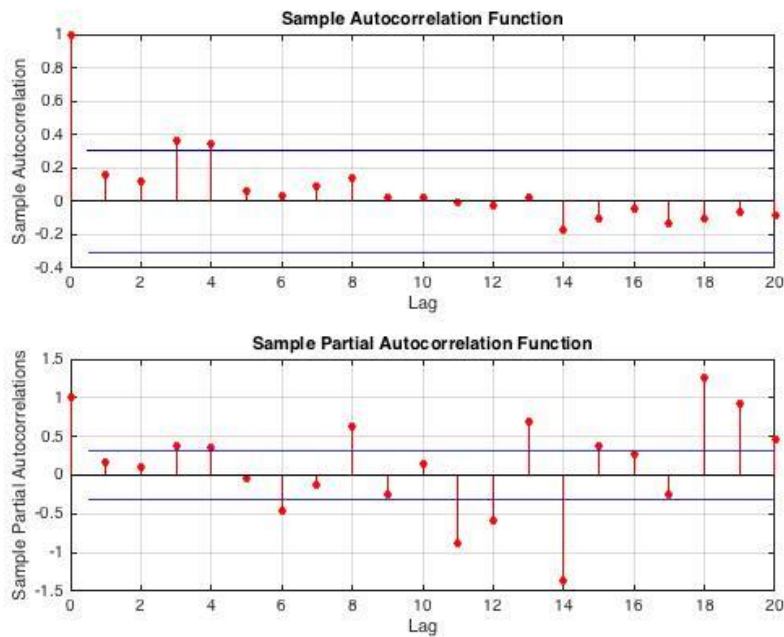


Figura 32 Model del camí aleatori BTV gràfic d'autocorrelacions

Es pot veure en tots els models que efectivament el gràfic de residuals es correspon amb un patró aleatori i està centrat pràcticament en el zero, afirmant que la diferenciació en les dades corregeix el patronatge dels residuals.

Pel que fa als gràfics d'autocorrelació es situen més dins dels límits esperats però encara amb alguna marcada tendència autoregressiva, com per exemple en el cas de València que encara hi és força present.

4.2.4. Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)

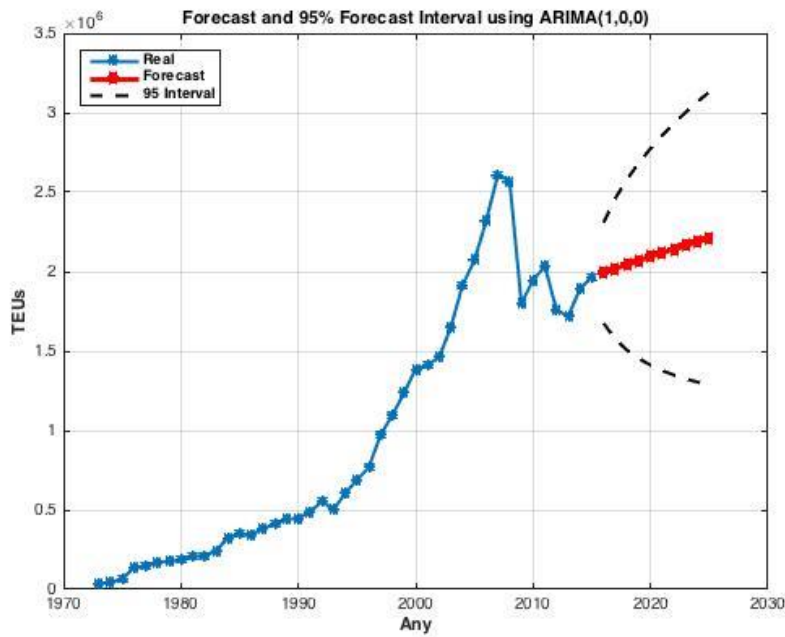


Figura 33 Model AR1 Barcelona

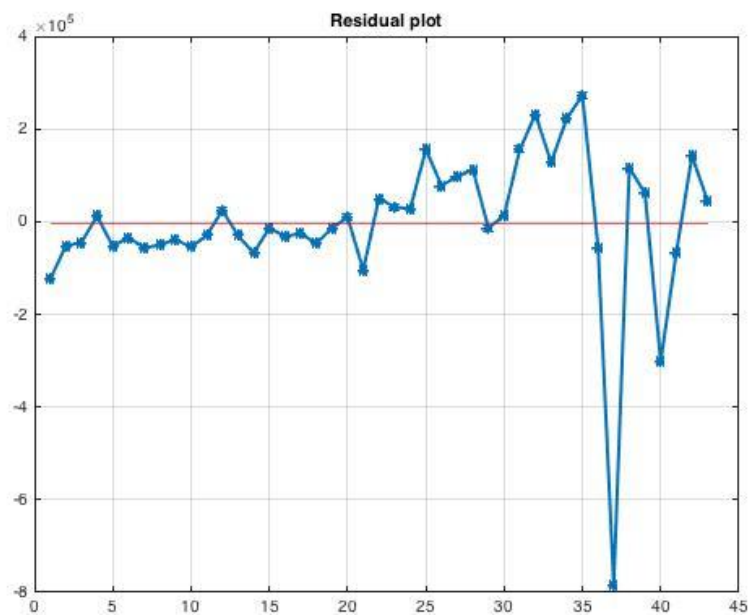


Figura 34 Model AR1 Barcelona gràfic de residuals

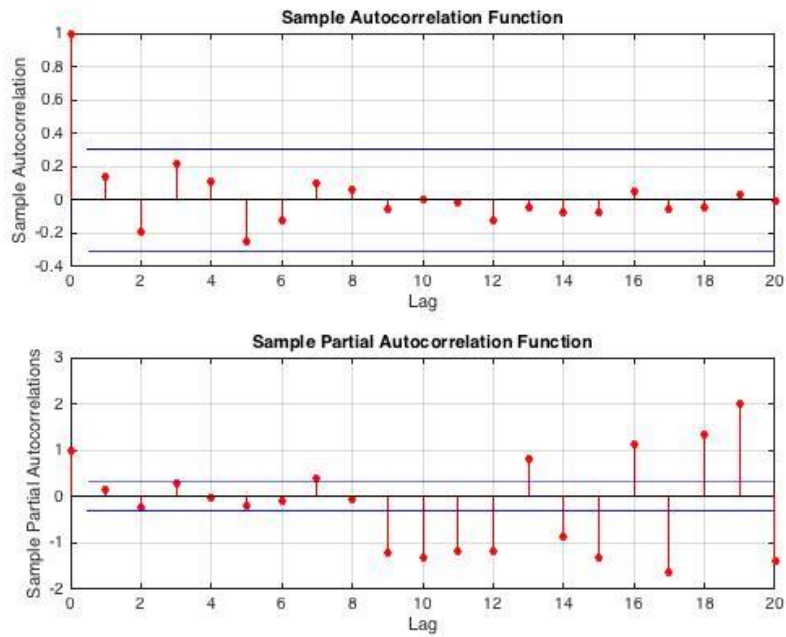


Figura 35 Model AR1 Barcelona gràfic d'autocorrelacions

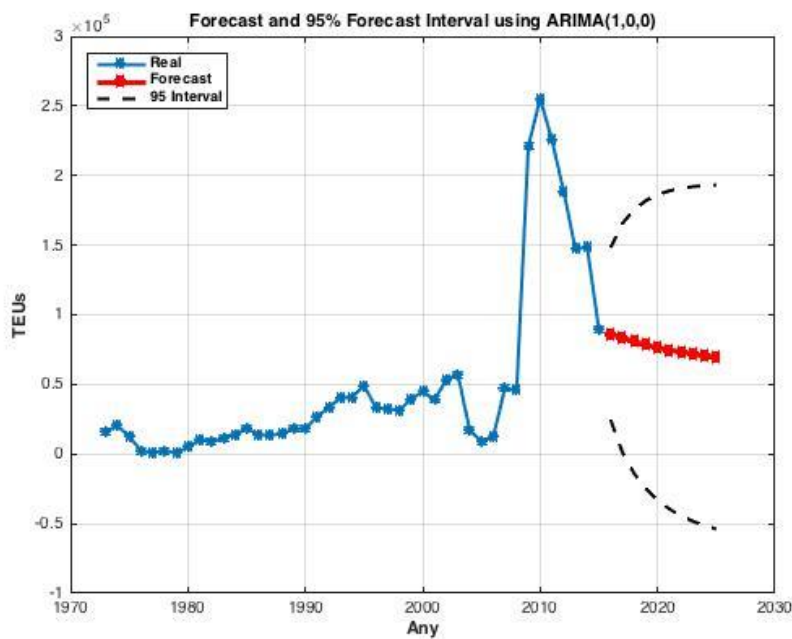


Figura 36 Model AR1 Tarragona

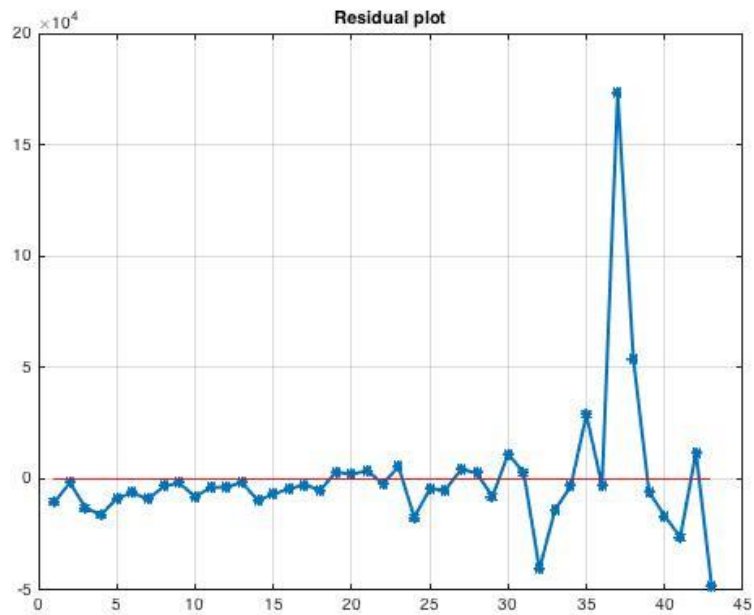


Figura 37 Model AR1 Tarragona gràfic de residuals

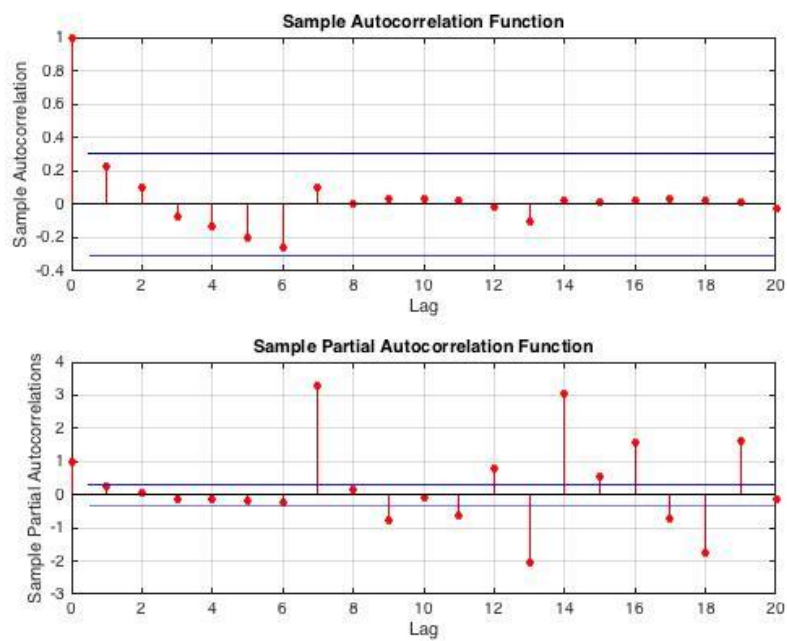


Figura 38 Model AR1 Tarragona gràfic d'autocorrelacions

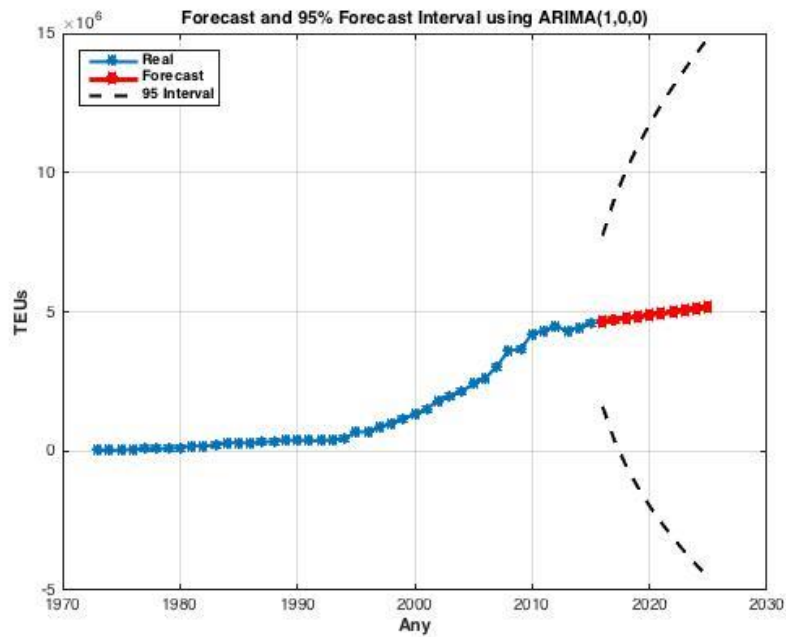


Figura 39 Model AR1 València

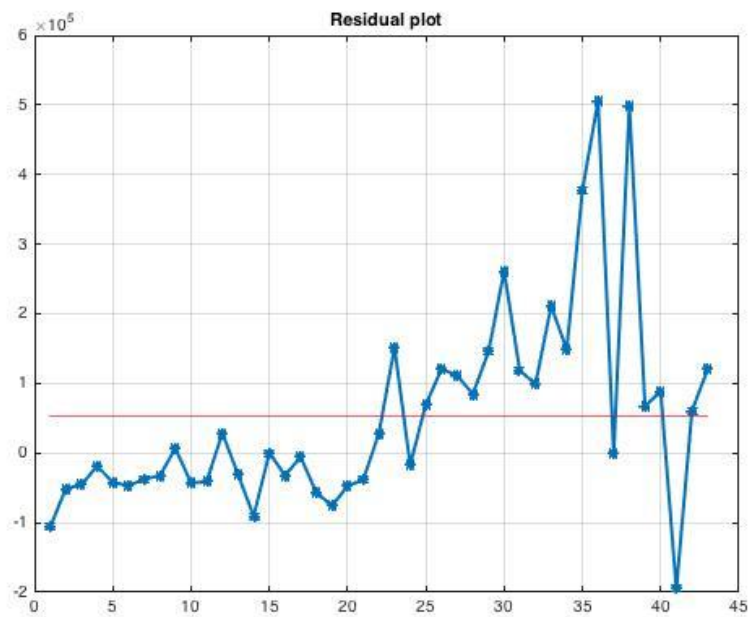


Figura 40 Model AR1 València gràfic de residualls

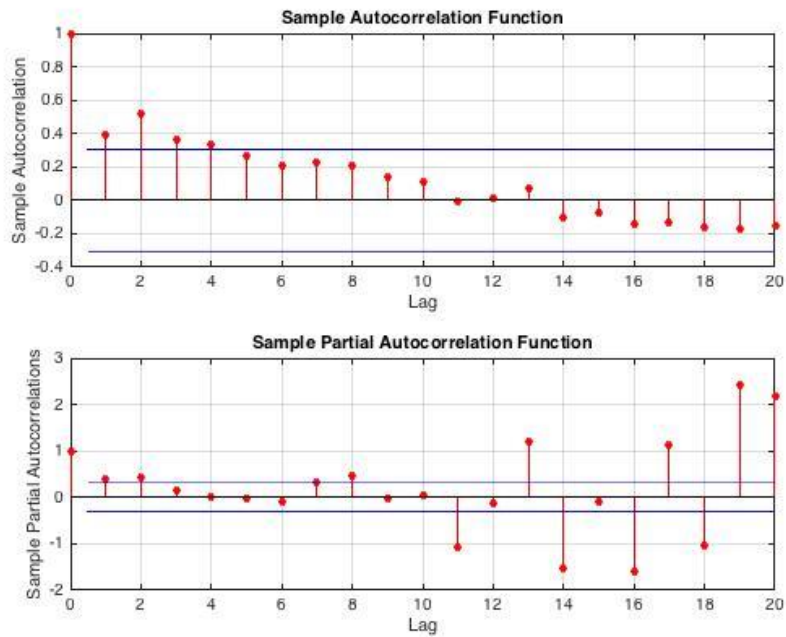


Figura 41 Model AR1 València gràfic d'autocorrelacions

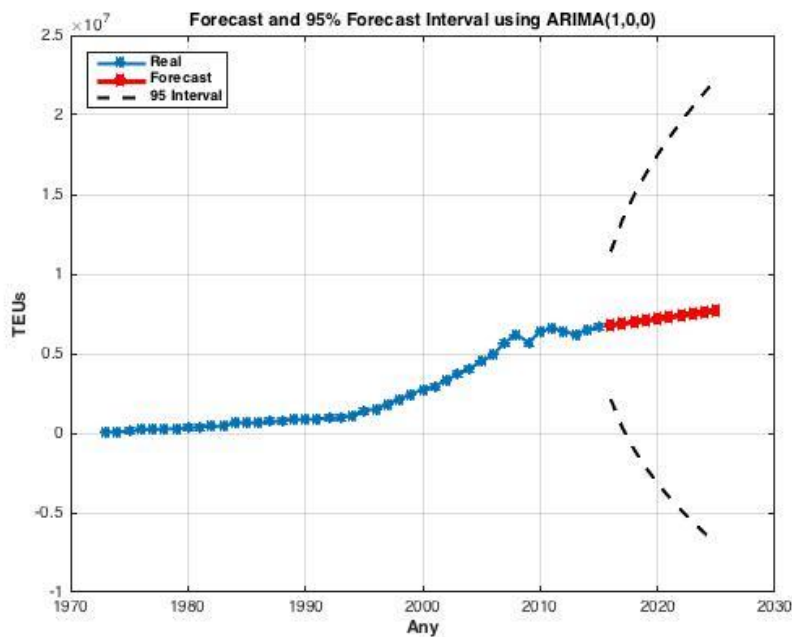


Figura 42 Model AR1 BTU

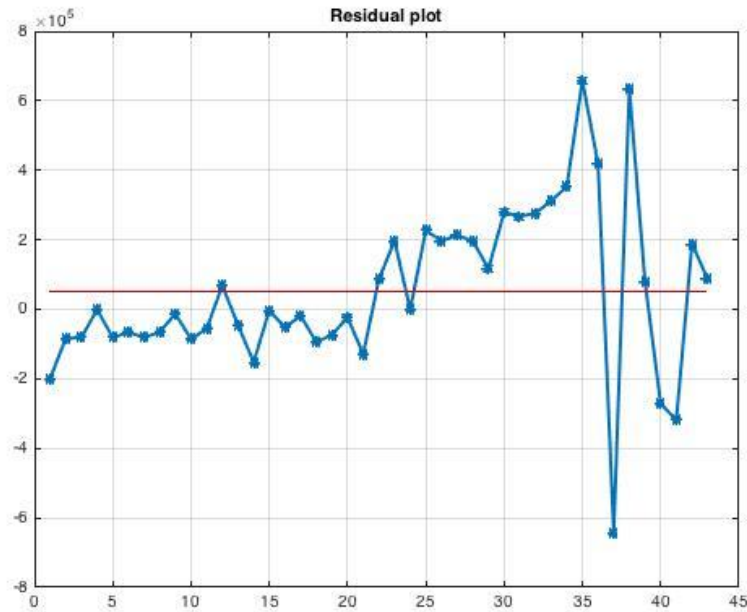


Figura 43 Model AR1 BTV gràfic de residuals

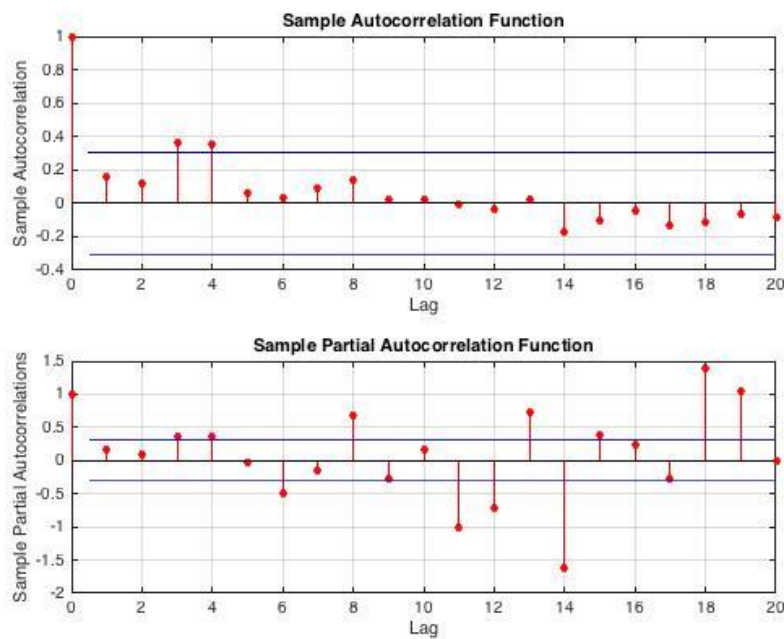


Figura 44 Model AR1 BTV gràfic d'autocorrelacions

Analitzant els gràfics es pot veure que en el cas dels residuals torna a haver-hi una distribució acceptable, però en els casos de València i els ports conjunts BTV aquests es centren a una certa distància respecte el zero teòric que interessa obtenir.

Pel que fa a les autocorrelacions en tots els casos s'ha corregit la forta tendència autoregressiva a excepció també de València, on s'hi troba encara les particularitats de les dades amb caràcter autoregressiu.

4.2.5. Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)

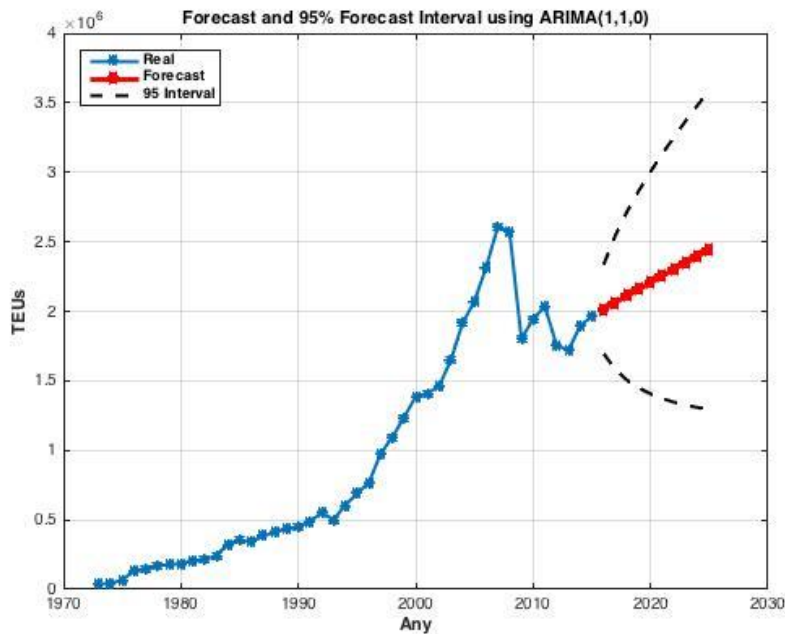


Figura 45 Model AR1 diferenciat Barcelona

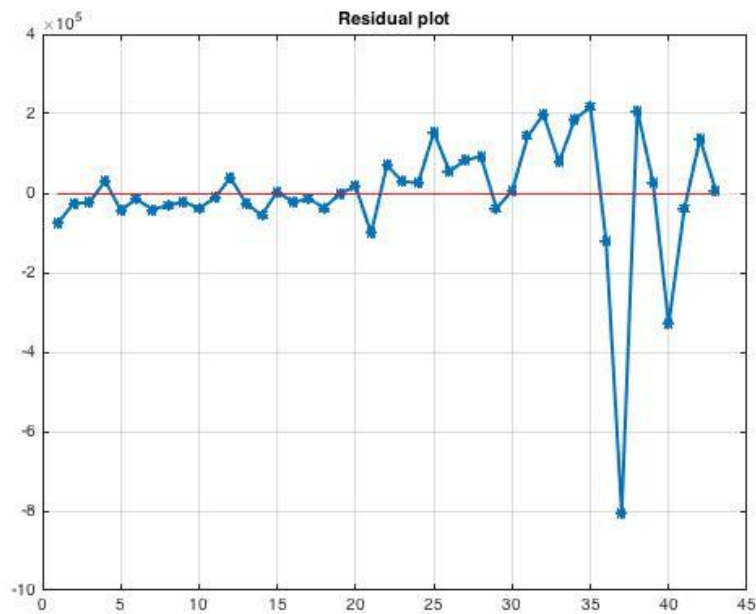


Figura 46 Model AR1 diferenciat Barcelona gràfic de residuals

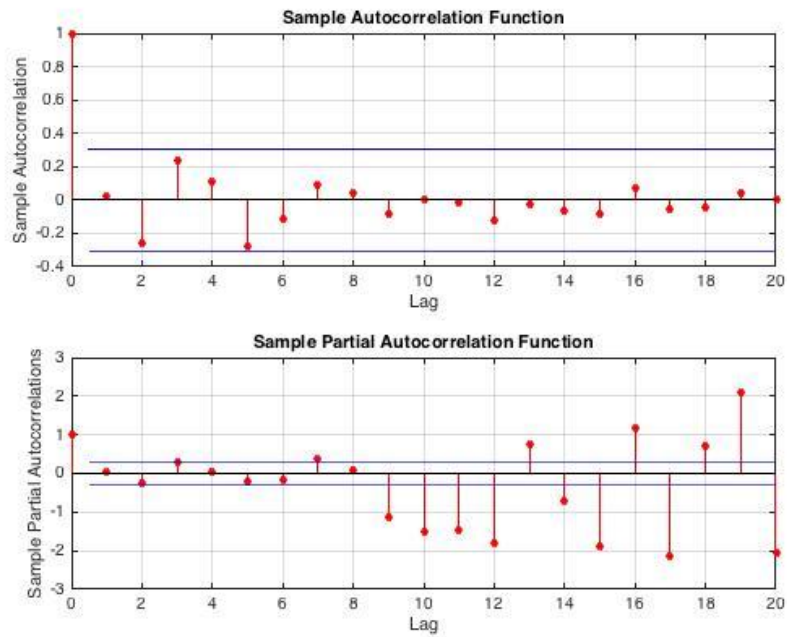


Figura 47 Model AR1 diferenciat Barcelona gràfic d'autocorrelacions

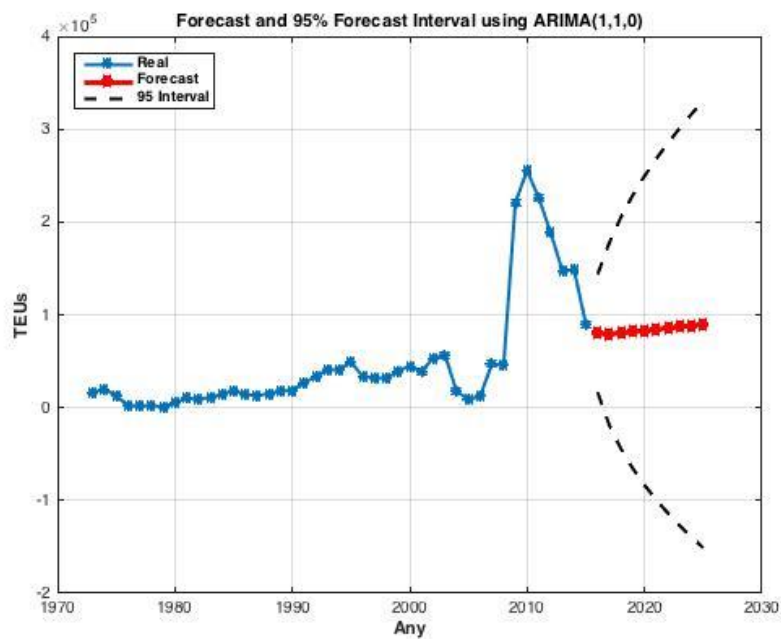


Figura 48 Model AR1 diferenciat Tarragona

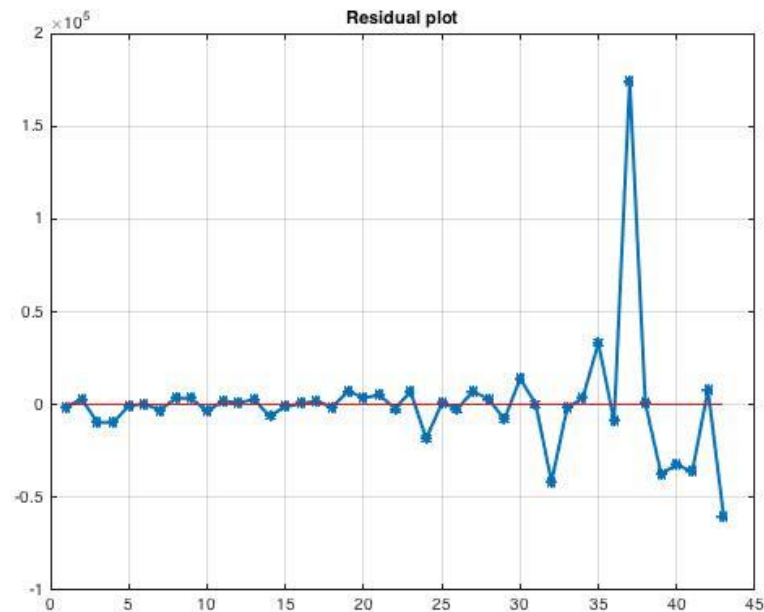


Figura 49 Model AR1 diferenciat Tarragona gràfic de residuals

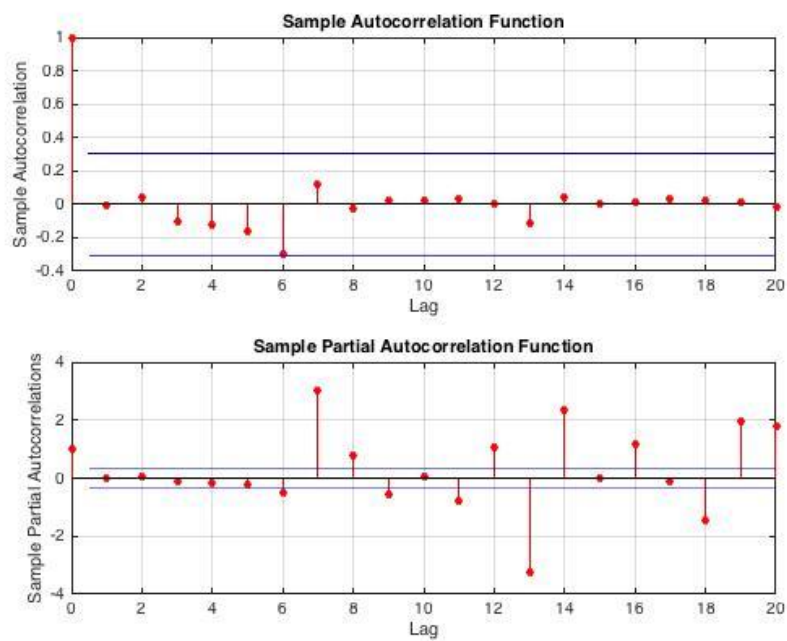


Figura 50 Model AR1 diferenciat Tarragona gràfic d'autocorrelacions

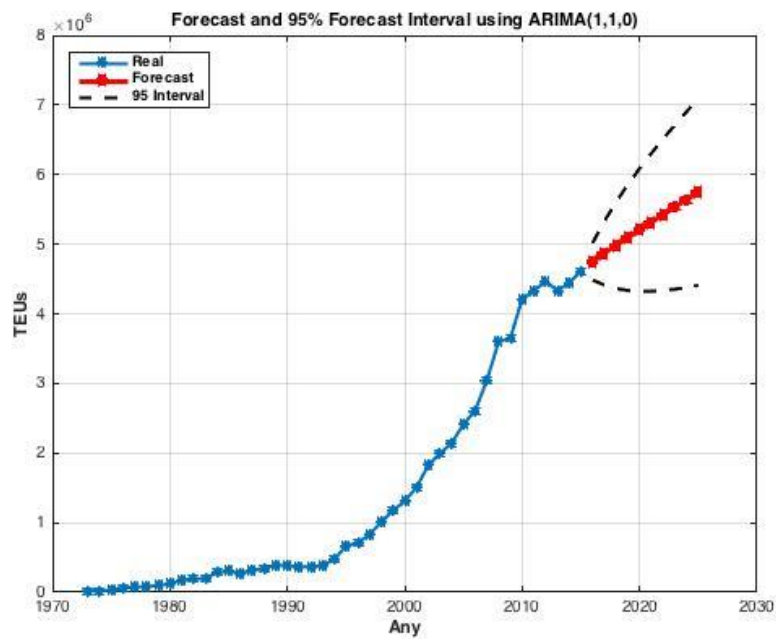


Figura 51 Model AR1 diferenciat València

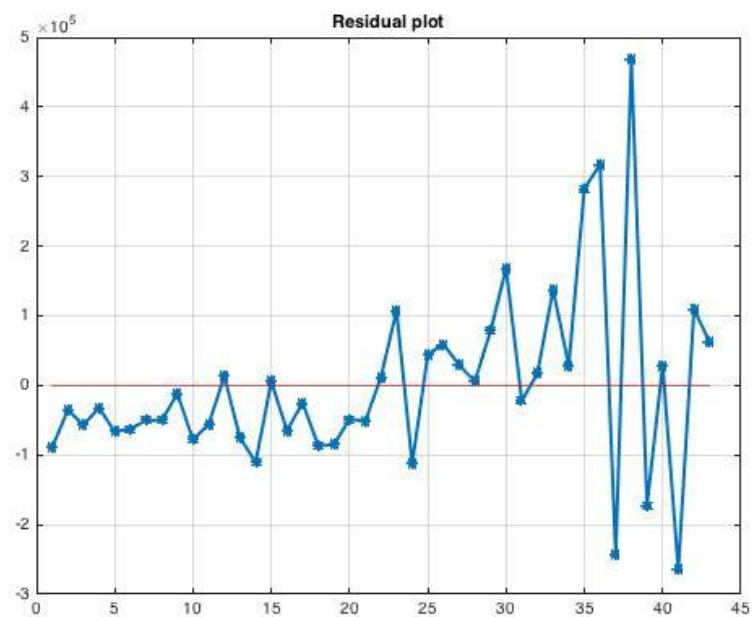


Figura 52 Model AR1 diferenciat València gràfic de residualls

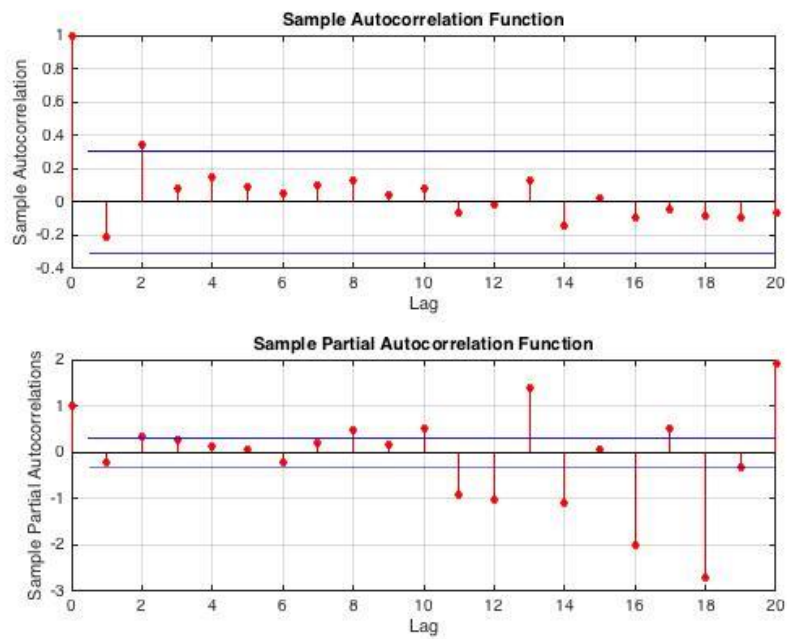


Figura 53 Model AR1 diferenciat València gràfic d'autocorrelacions

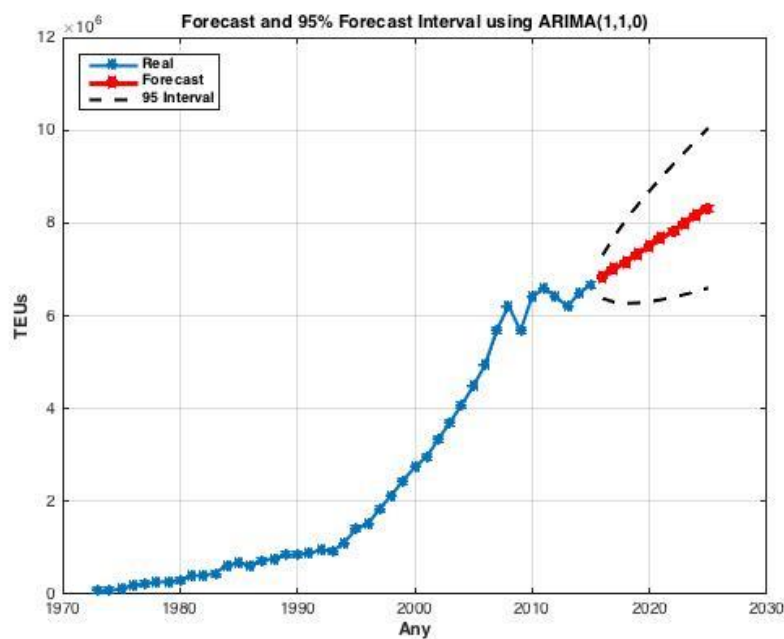


Figura 54 Model AR1 diferenciat BTU

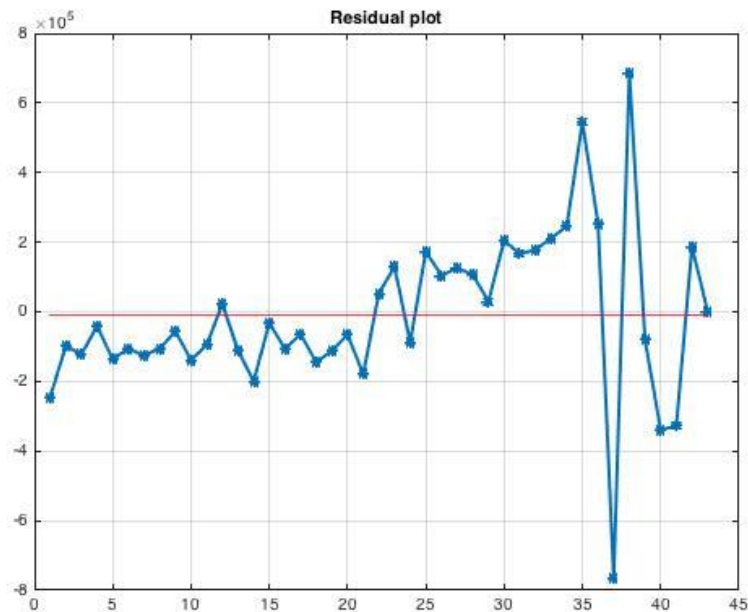


Figura 55 Model AR1 diferenciat BTV gràfic de residuals

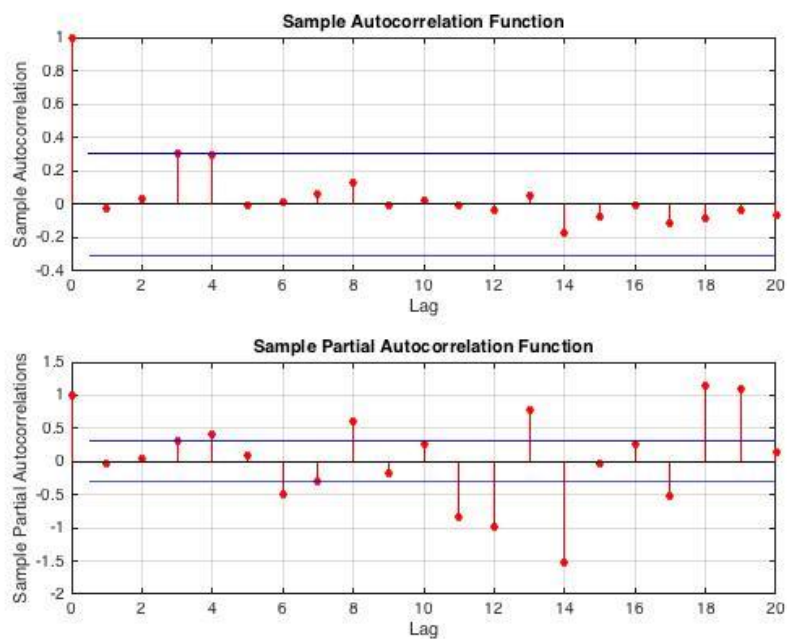


Figura 56 Model AR1 diferenciat BTV gràfic d'autocorrelacions

En tots els models s'observa uns bons gràfics pel que fa a residuals i autocorrelació, per tant, si és cert que en aquest cas s'equilibra en tots els casos aquests gràfics i poden ser acceptats com a correctes.

4.2.6. Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)

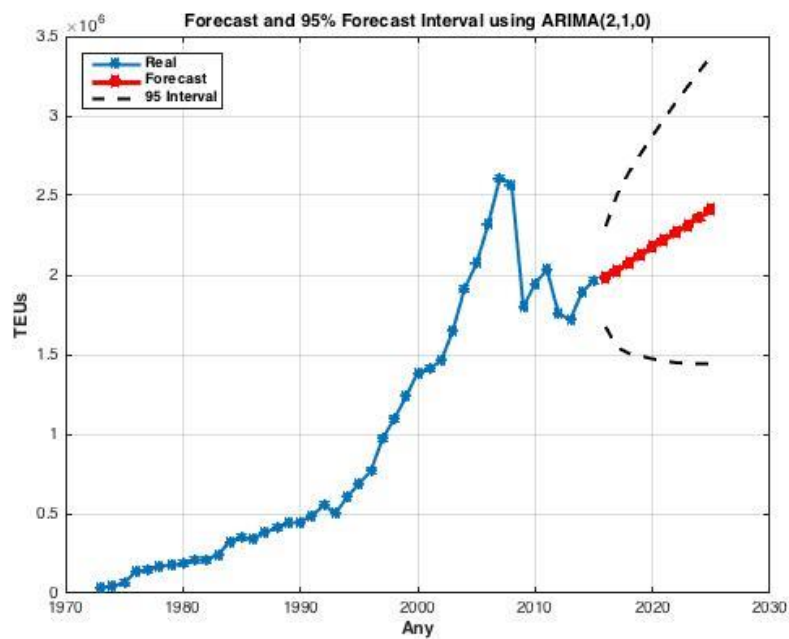


Figura 57 Model AR2 diferenciat Barcelona

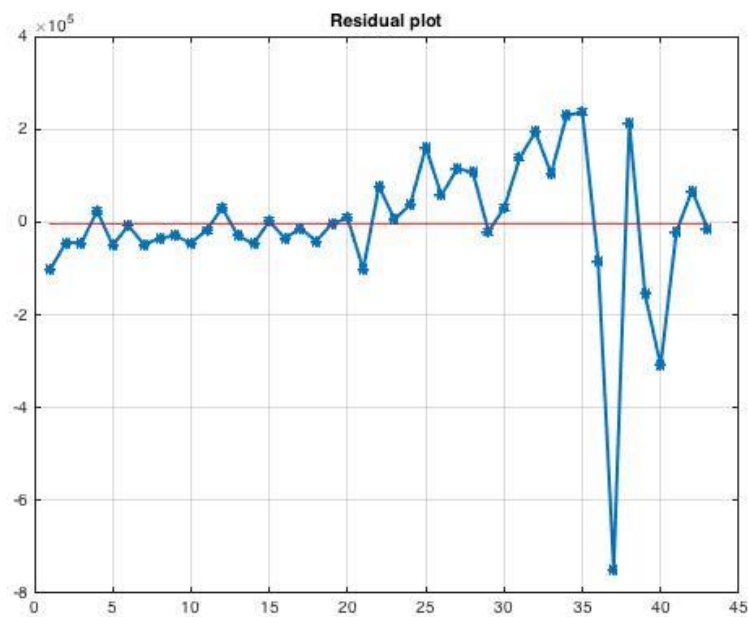


Figura 58 Model AR2 diferenciat Barcelona gràfic de residuals

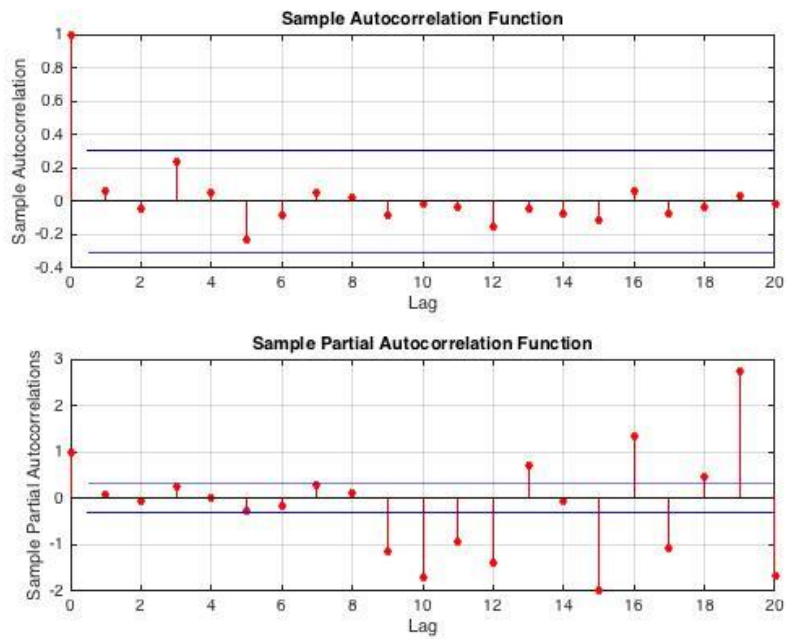


Figura 59 Model AR2 diferenciat Barcelona gràfic d'autocorrelacions

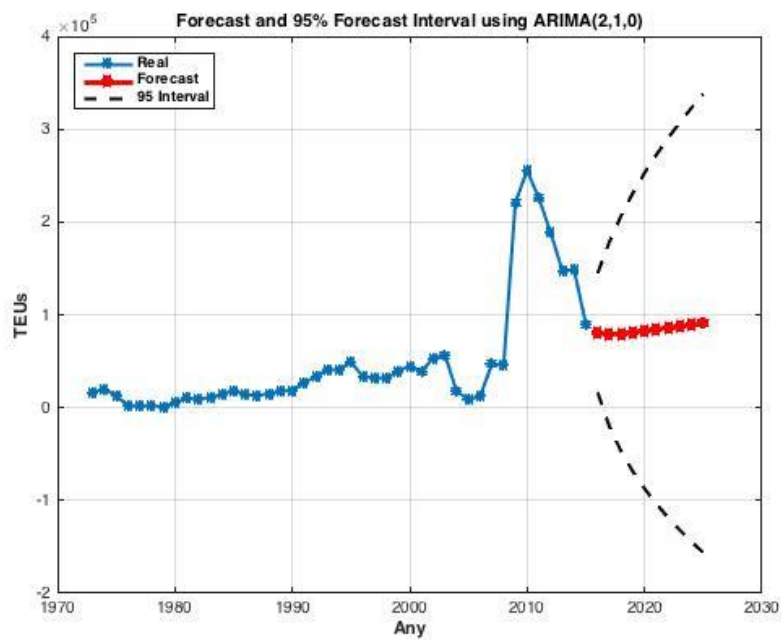


Figura 60 Model AR2 diferenciat Tarragona

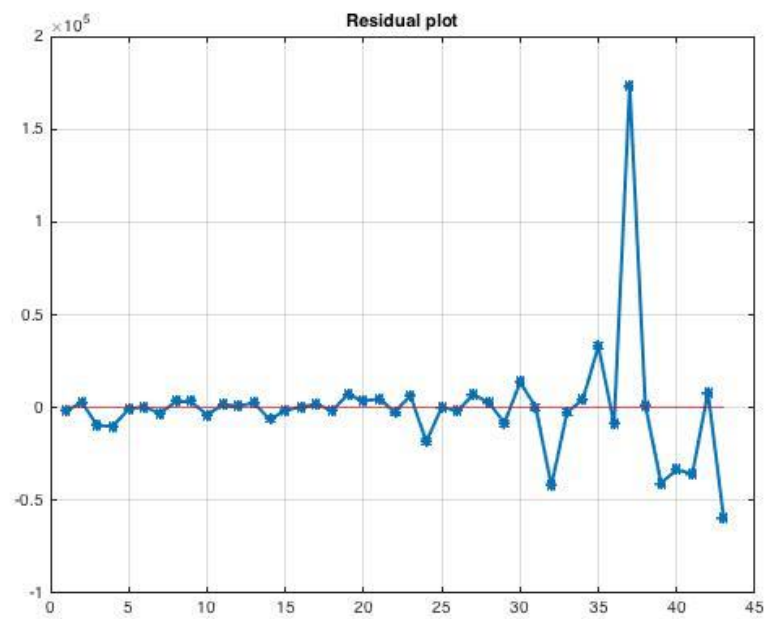


Figura 61 Model AR2 diferenciat Tarragona gràfic de residuals

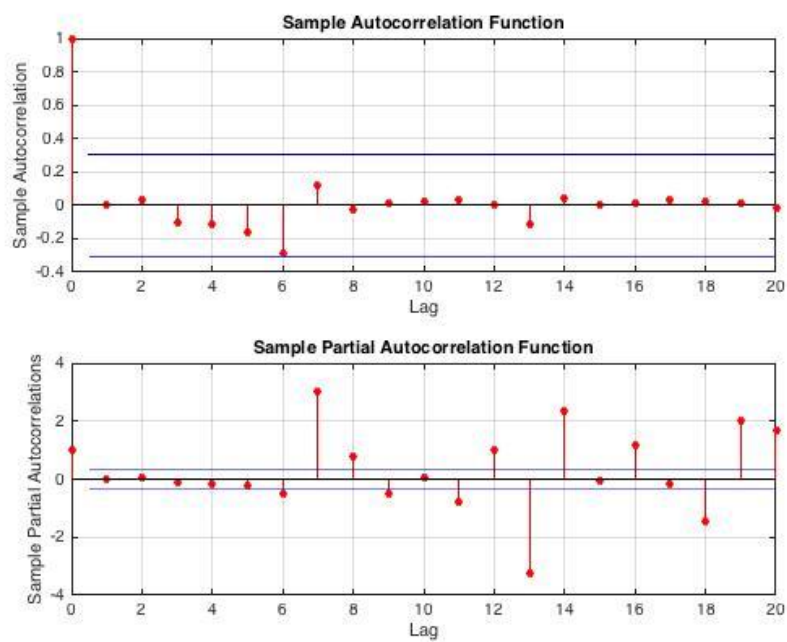


Figura 62 Model AR2 diferenciat Tarragona gràfic d'autocorrelacions

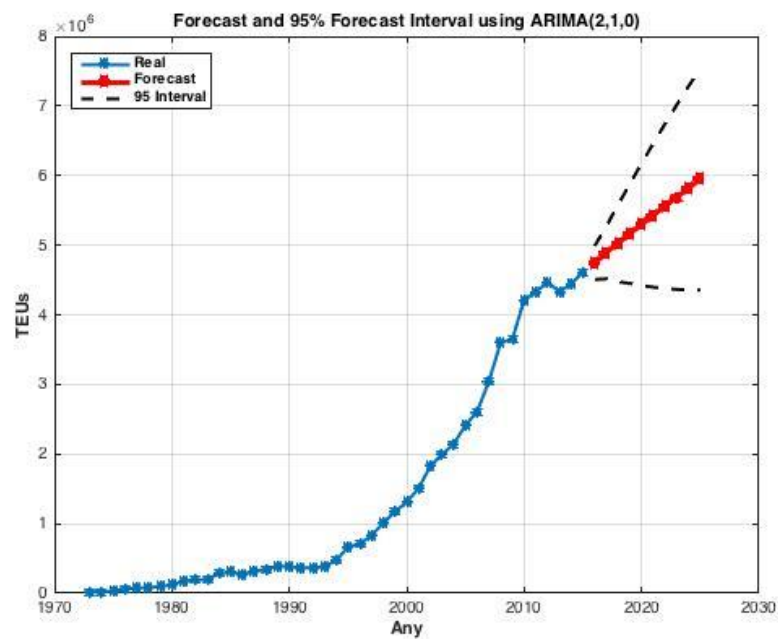


Figura 63 Model AR2 diferenciat València

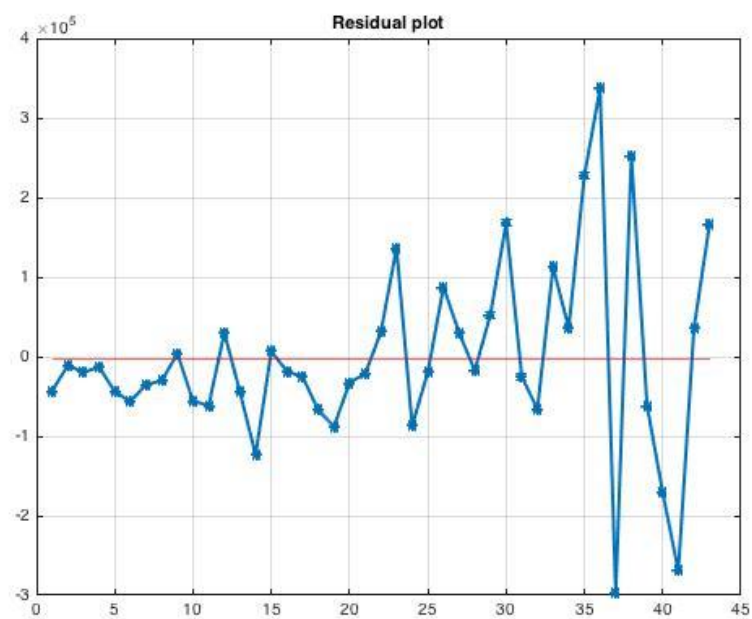


Figura 64 Model AR2 diferenciat València gràfic de residuals

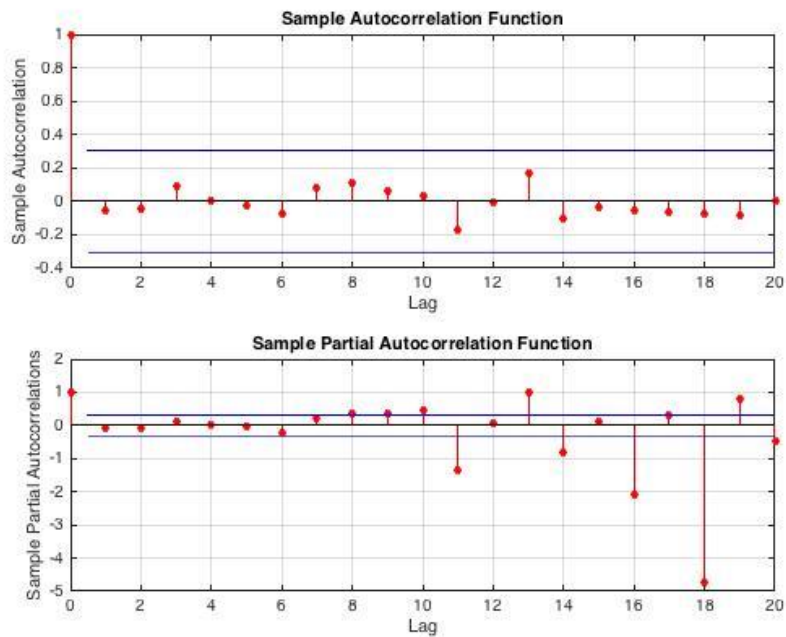


Figura 65 Model AR2 diferenciat València gràfic d'autocorrelacions

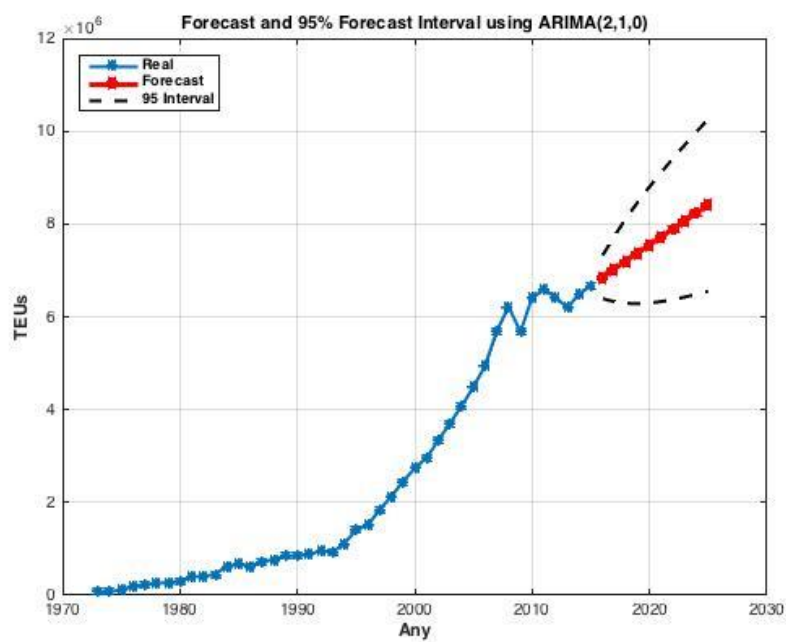


Figura 66 Model AR2 diferenciat BTV

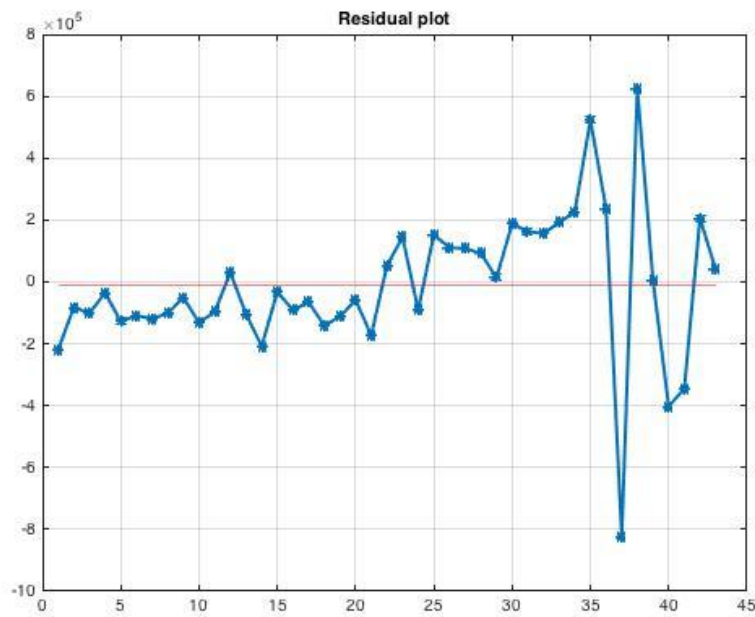


Figura 67 Model AR2 diferenciat BTV gràfic de residuals

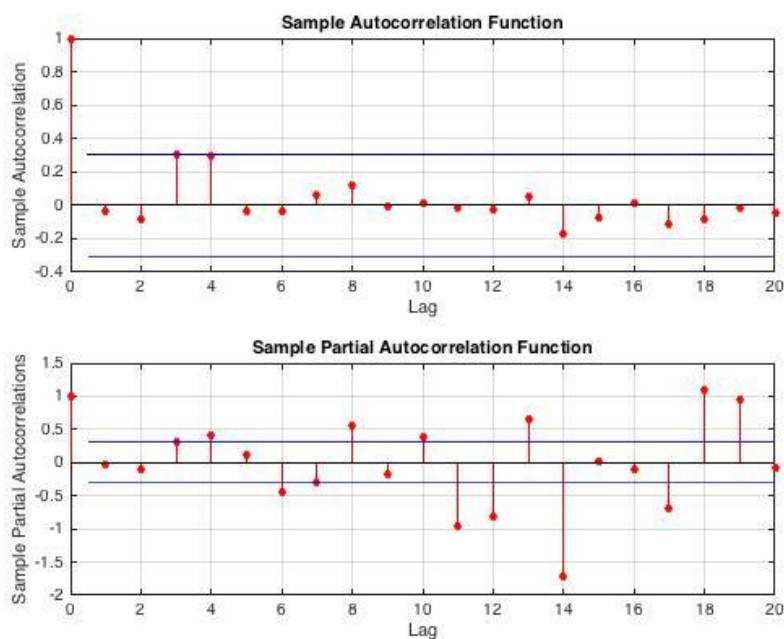


Figura 68 Model AR2 diferenciat BTV gràfic d'autocorrelacions

El model funciona bé a nivell teòric com es pot veure en els diferents gràfics, però no hi aporta una millora clara respecte qualsevol dels models anteriors, que guanyen en simplicitat.

4.2.7. Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)

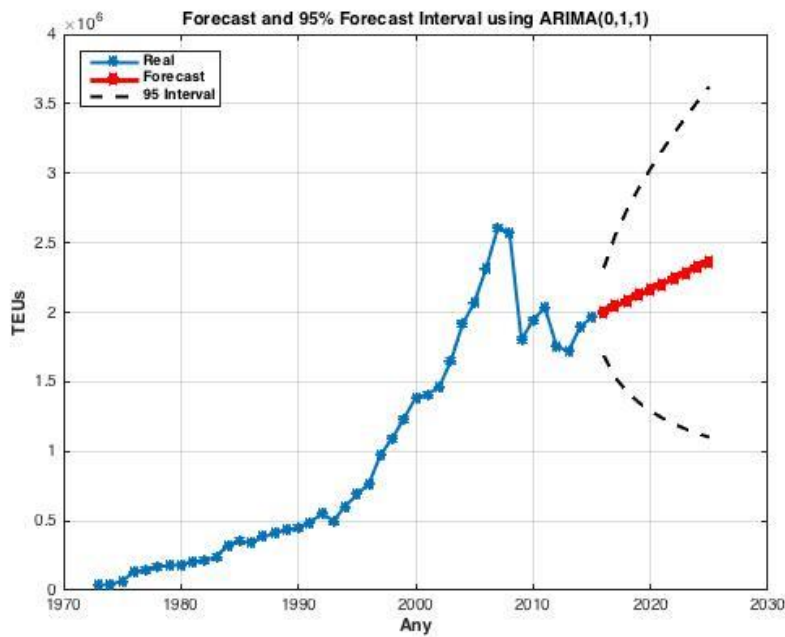


Figura 69 Model d'ajustament exponencial Barcelona

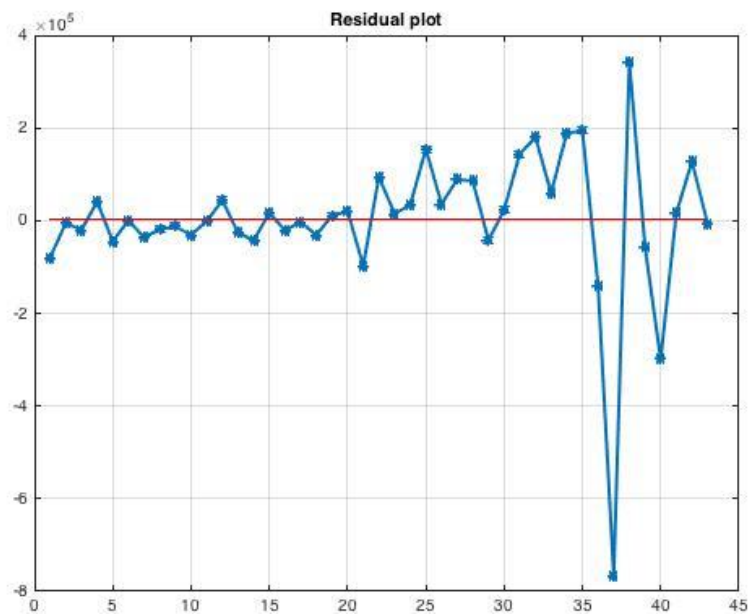


Figura 70 Model d'ajustament exponencial Barcelona gràfic de residuals

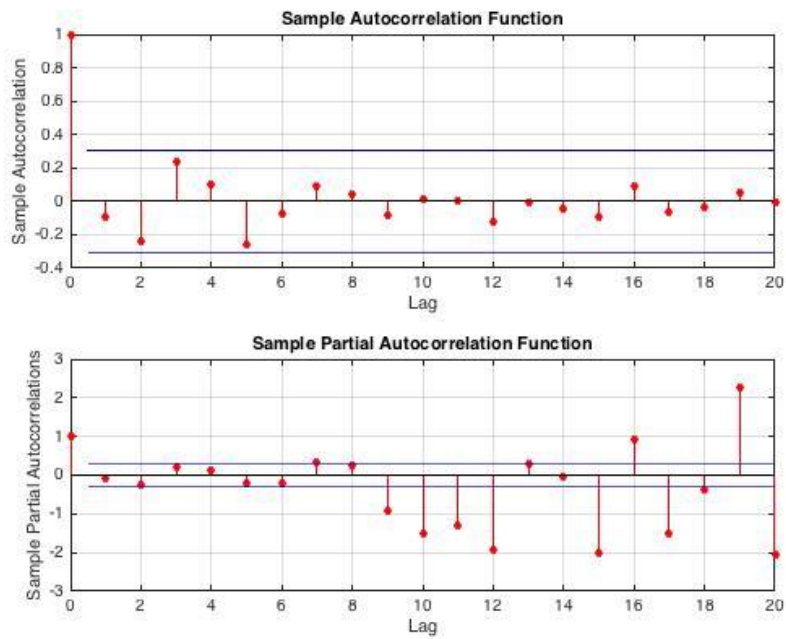


Figura 71 Model d'ajustament exponencial Barcelona gràfic d'autocorrelacions

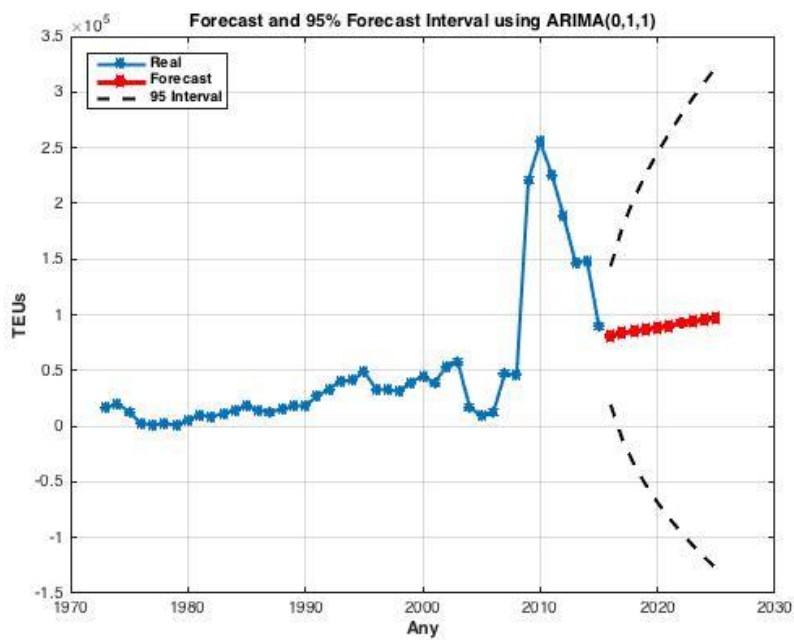


Figura 72 Model d'ajustament exponencial Tarragona

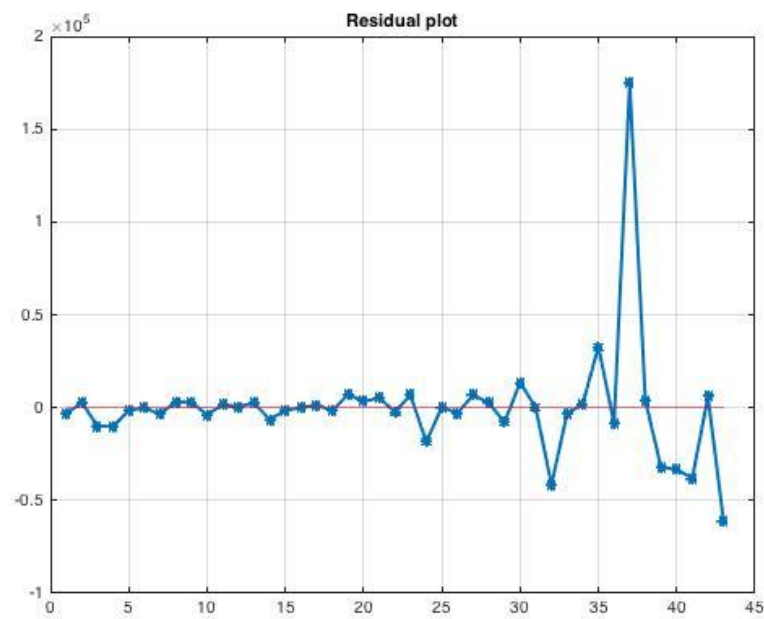


Figura 73 Model d'ajustament exponencial Tarragona gràfic de residuals

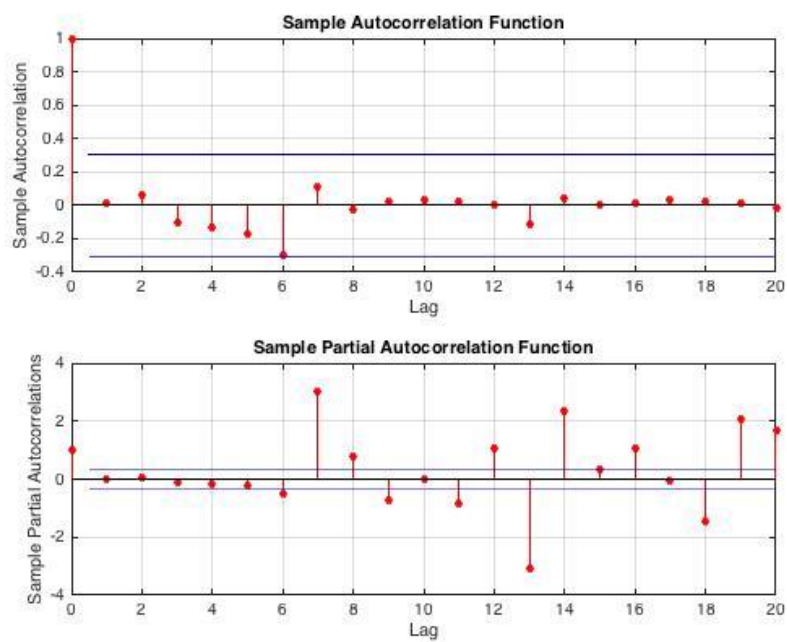


Figura 74 Model d'ajustament exponencial Tarragona gràfic d'autocorrelacions

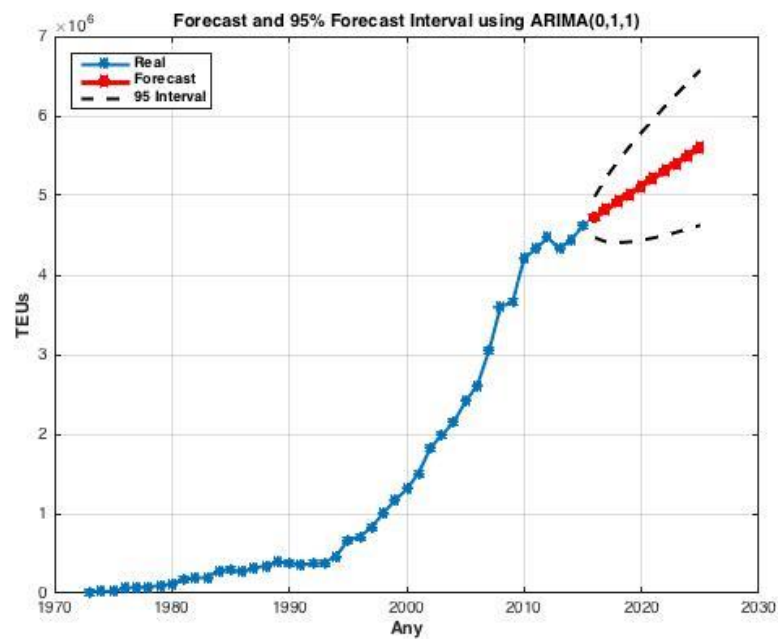


Figura 75 Model d'ajustament exponencial València

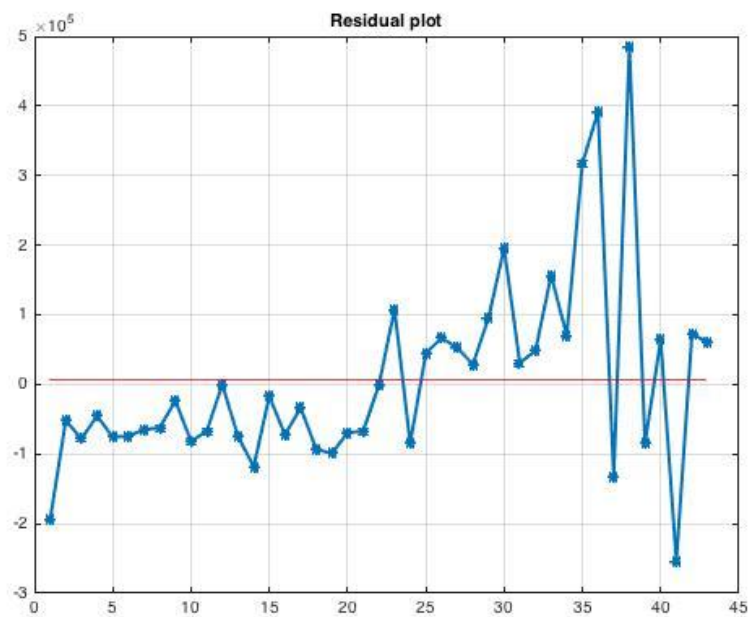


Figura 76 Model d'ajustament exponencial València gràfic de residuals

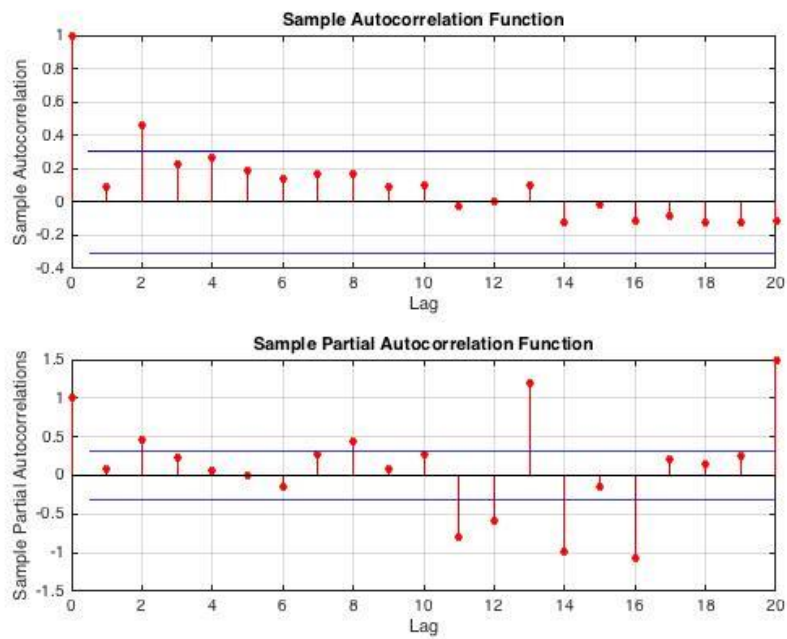


Figura 77 Model d'ajustament exponencial València gràfic d'autocorrelacions

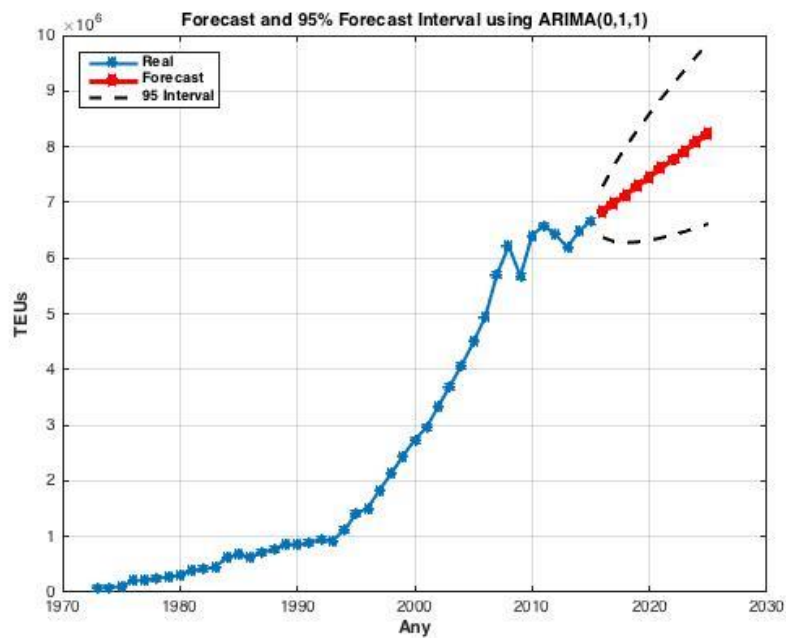


Figura 78 Model d'ajustament exponencial BTV

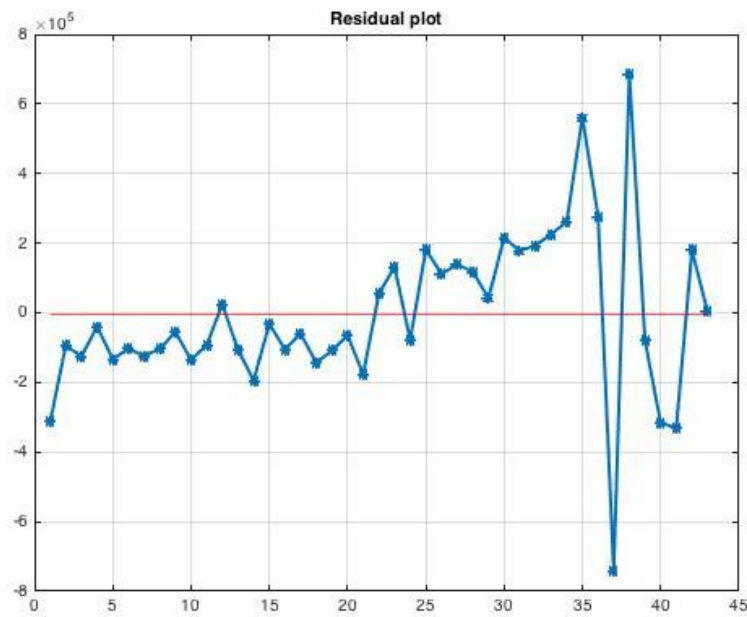


Figura 79 Model d'ajustament exponencial BTV gràfic de residuals

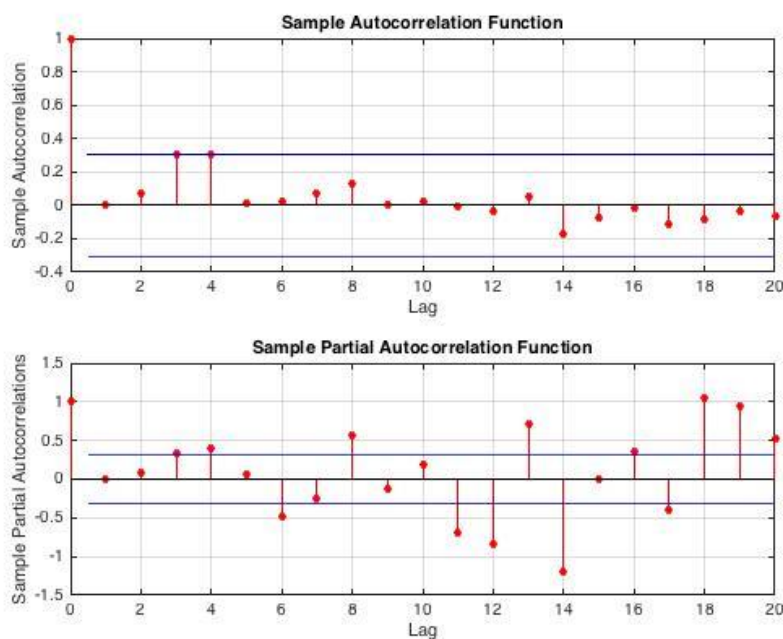


Figura 80 Model d'ajustament exponencial BTV gràfic d'autocorrelacions

En aquest model els residuals segueixen un patró acceptable, però en el gràfic d'autocorrelacions hi ha variacions que si bé es cert que són en *lags* avançats, hi havia altres models en que estaven més ajustats.

En d'altres casos com en el de València, els valors eren acceptables però l'autocorrelació positiva porta a pensar que l'ajustament amb un terme autoregressiu hi funciona millor.

El fet que en certs *lags* existeixi aquesta variació puntual molt gran s'explica de la teoria mateixa d'un model de mitjana mòbil. Tot i que els valors presents tenen més pes pel propi ajust exponencial del model, la variació en el creixement dels TEU sobretot en el conjunt de període abans (creixement marcat), durant (estancament i en certs casos davallada) i després de la crisi, afecta notablement en l'ajust de les mitjanes. Això sumat a la forta tendència autoregressiva observada des dels primers models, fa que tot i que per norma general aquest tipus de model sigui un bon candidat a model vàlid, perdi força davant dels altres.

4.3.EXERCICIS DE VERIFICACIÓ 2010 I 2005

Els resultats en aquest apartat es presentaran diferenciant cadascun dels models, i dins de cadascun es presentaran primer els gràfics del 2010 i després del 2005. A l'acabar cada model es faran alguns apunts dels gràfics observats, que com en l'apartat anterior ajudaran a extreure posteriorment idees sobre el funcionament que ha tingut cadascun dels models estudiats.

4.3.1. Regressió simple

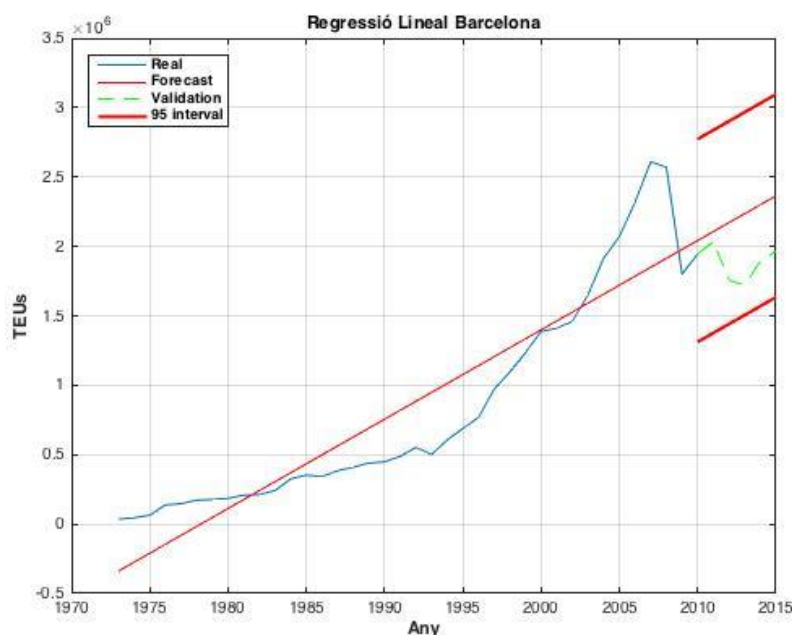


Figura 81 Regressió simple Barcelona verificació 2010

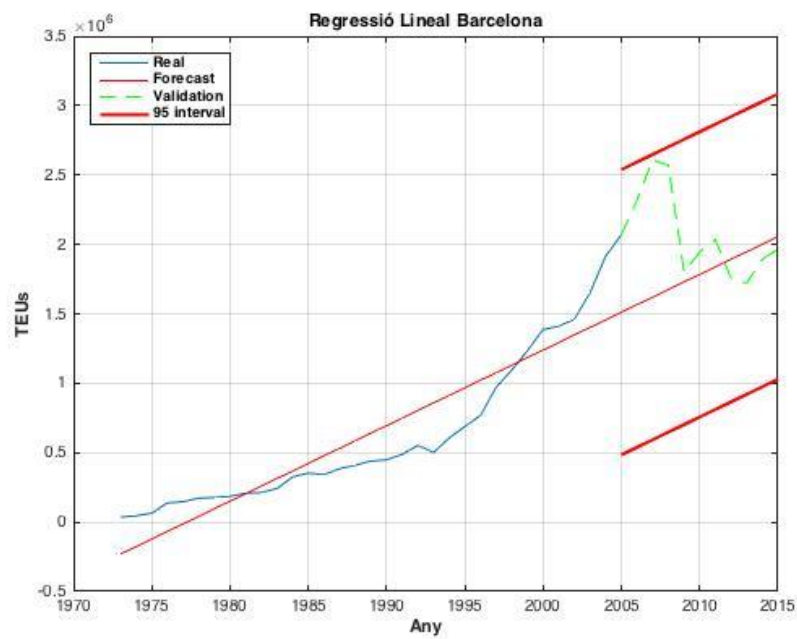


Figura 82 Regressió simple Barcelona verificació 2005

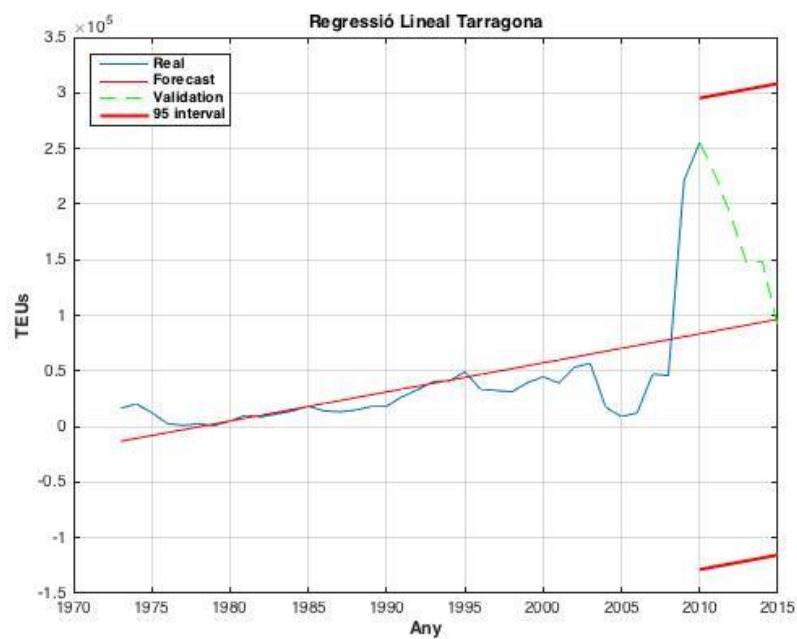


Figura 83 Regressió simple Tarragona verificació 2010

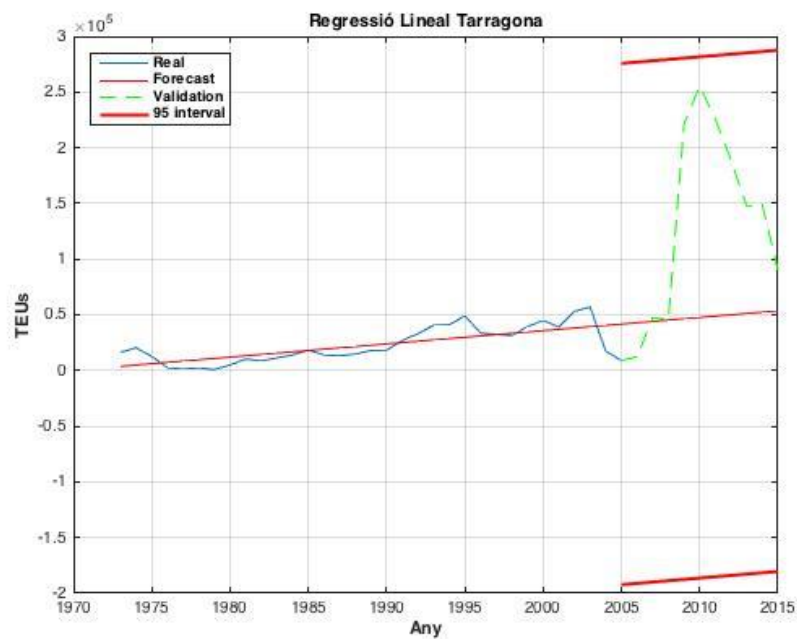


Figura 84 Regressió simple Tarragona verificació 2005

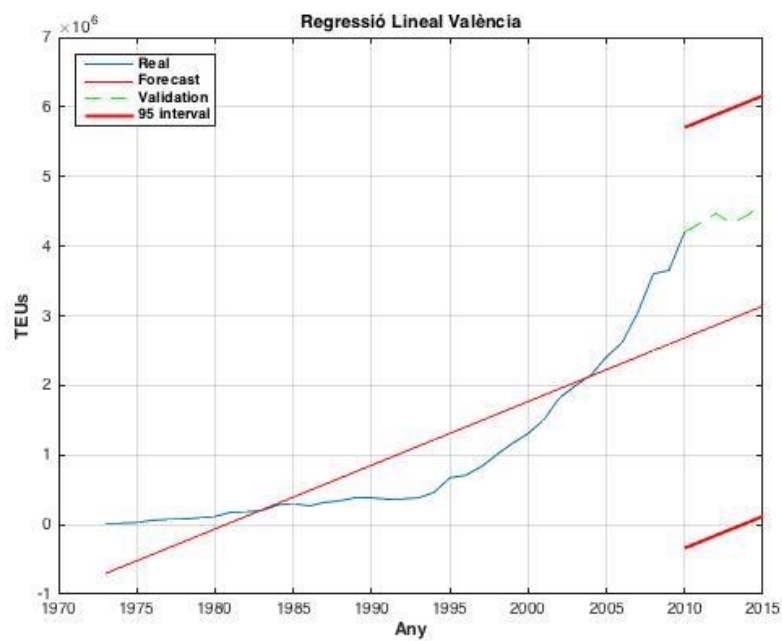


Figura 85 Regressió simple València verificació 2010

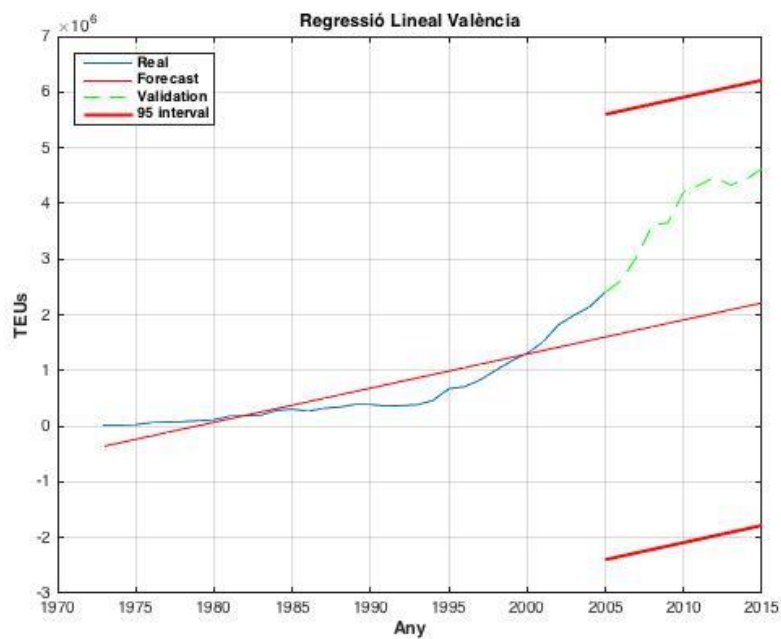


Figura 86 Regressió simple València verificació 2005

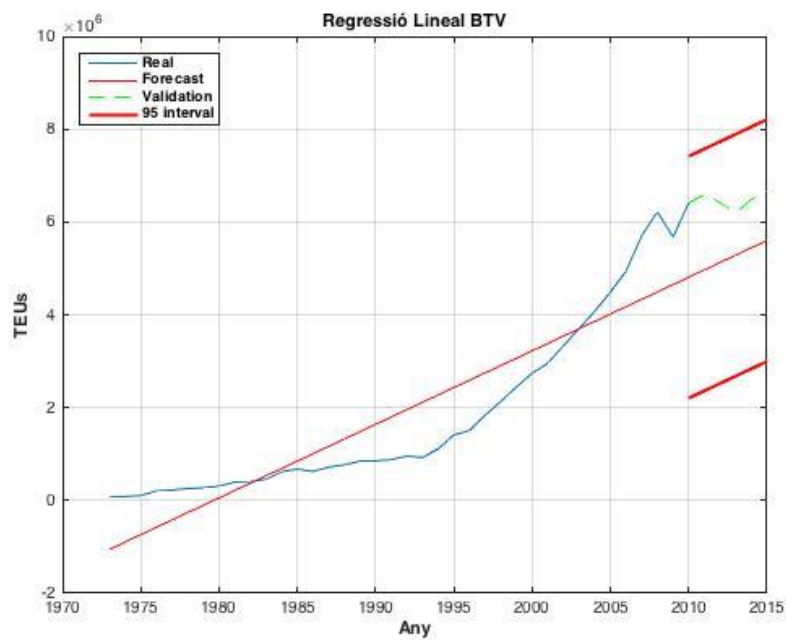


Figura 87 Regressió simple BTV verificació 2010

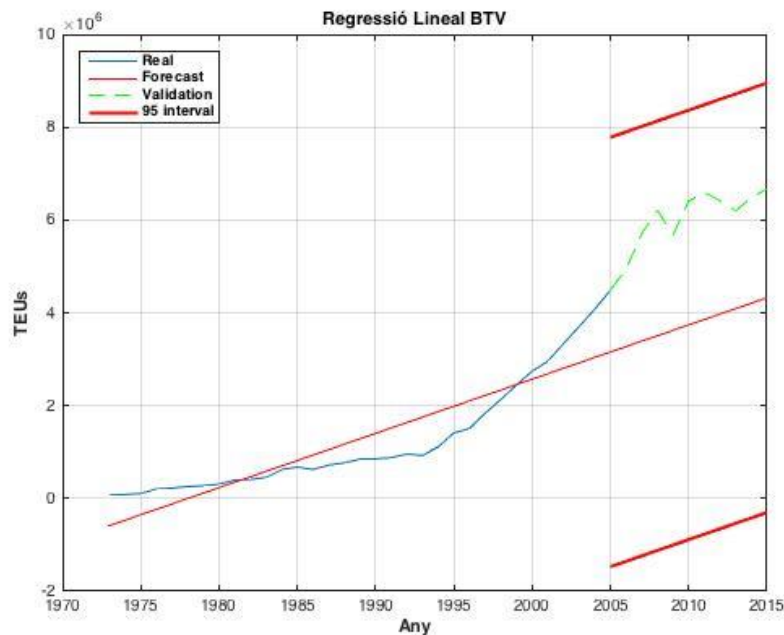


Figura 88 Regressió simple BTV verificació 2005

En tots els casos s'observa que el model dona una tendència clarament cap a augmentar els TEU amb el temps, i també en tots els casos en els dos períodes analitzats la realitat s'ha trobat dins dels intervals de confiança. S'hi observa que per aquest model els intervals de confiança són lineals, i en general han estat més amples en el període de verificació del 2005 que en el del 2010, indicant la idea que com més lluny volem predir, més incertesa tindrem. Aquest concepte de forma teòrica s'hauria d'anar repetint en els diferents casos.

4.3.2. Model de la mitjana (ARIMA(0,0,0)+c)

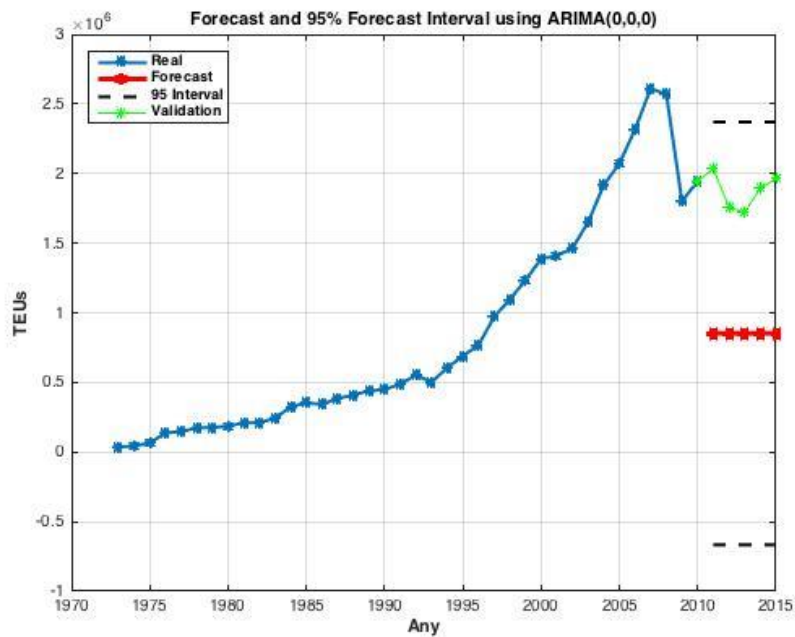


Figura 89 Model de la mitjana Barcelona verificació 2010

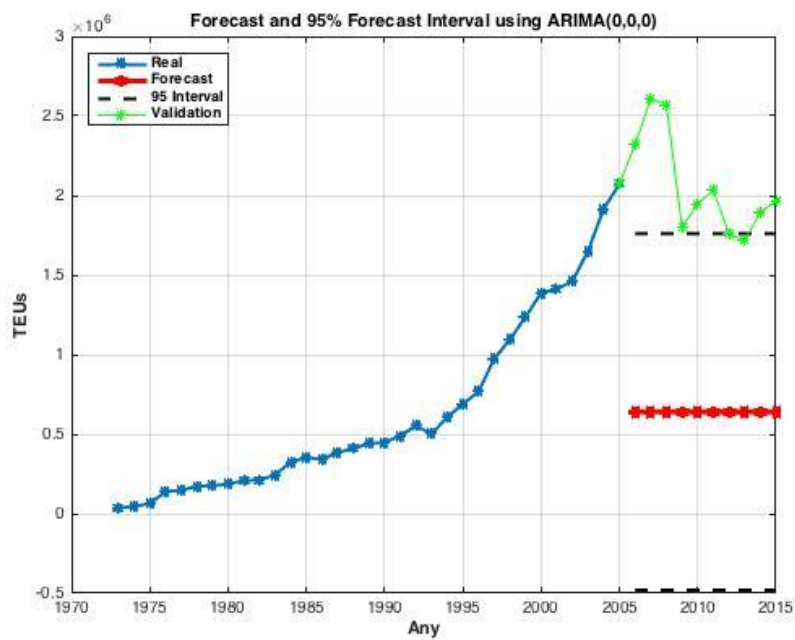


Figura 90 Model de la mitjana Barcelona verificació 2005

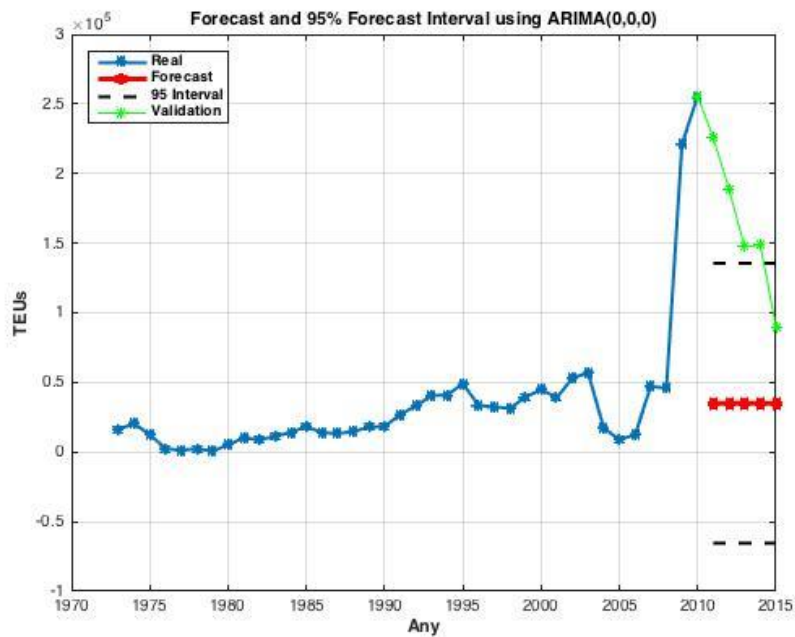


Figura 91 Model de la mitjana Tarragona verificació 2010

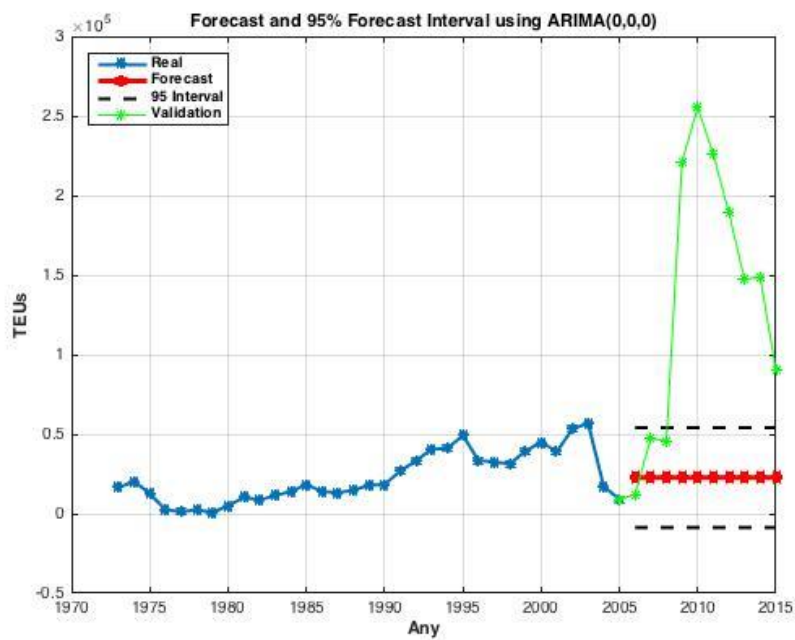


Figura 92 Model de la mitjana Tarragona verificació 2005

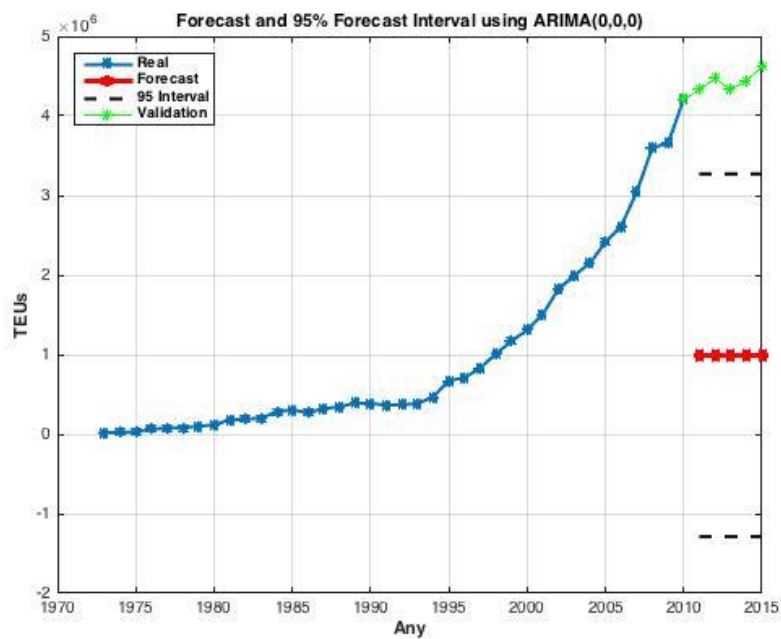


Figura 93 Model de la mitjana València verificació 2010

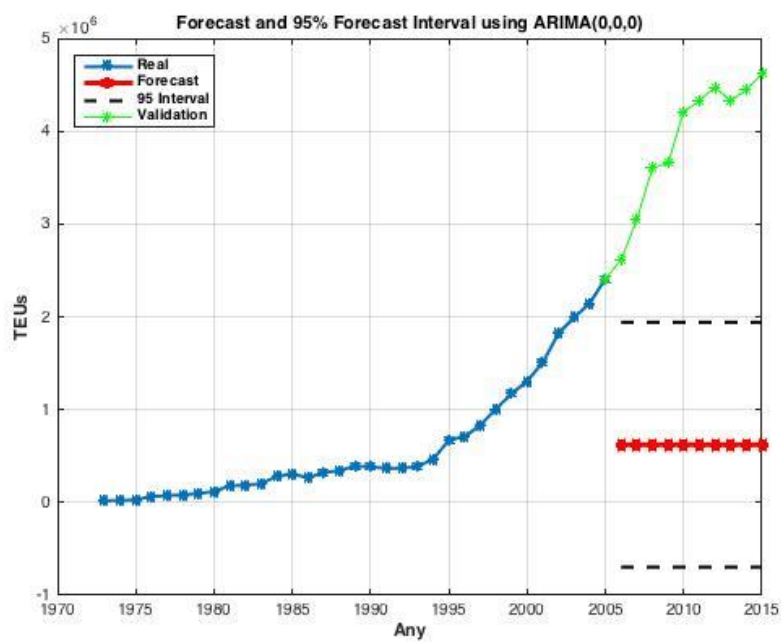


Figura 94 Model de la mitjana València verificació 2005

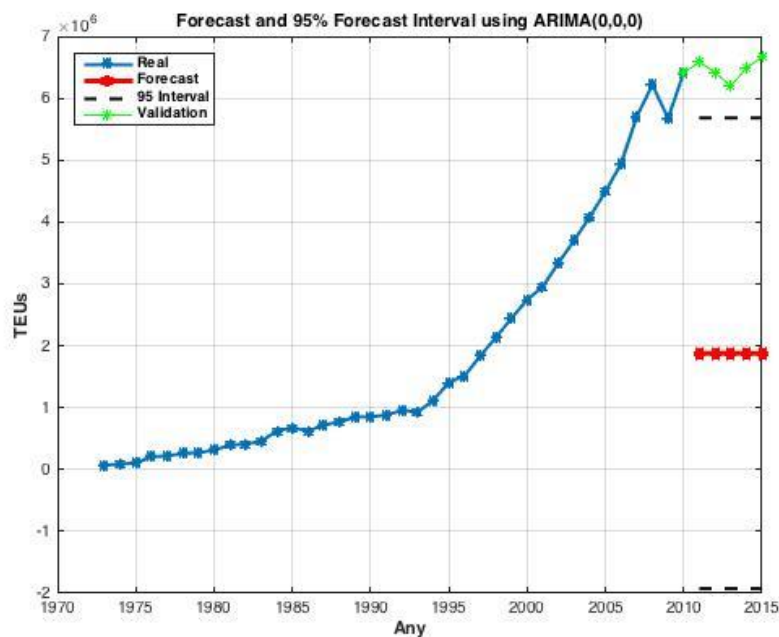


Figura 95 Model de la mitjana BTV verificació 2010

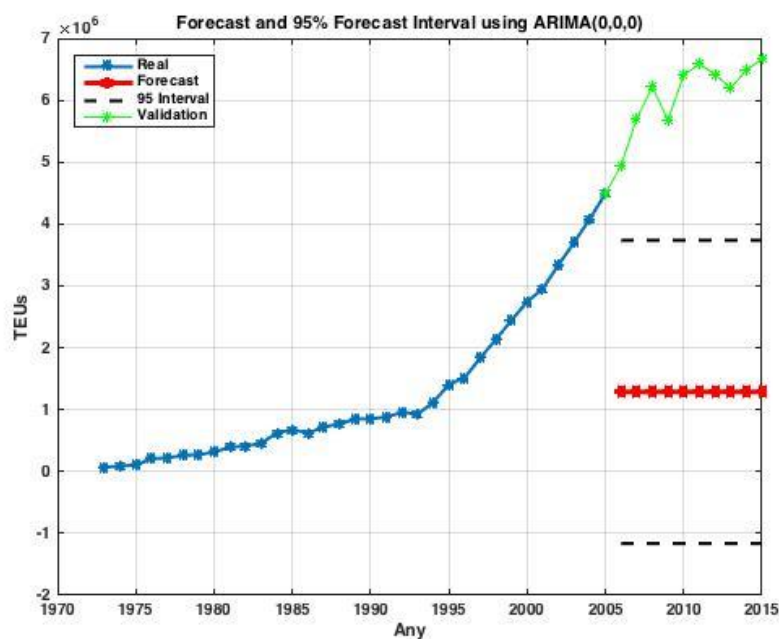


Figura 96 Model de la mitjana BTV verificació 2005

La predicció d'aquest model és puntual, per tant, d'entrada és difícil que serveixi amb les dades veient la seva naturalesa a tendir a l'alça. També es pot veure que els intervals de confiança en general són molt amples, però tot i això en molts dels períodes de verificació la realitat ha quedat per fora d'aquest, per tant no serviria el model.

El cas de Tarragona és clarament el més diferenciat i s'hi veu el que s'ha esmentat anteriorment pel que fa als models en la predicció de grans canvis de tendència, sobretot en la verificació del 2005, on el model indica que tot seguirà amb una tendència similar quan en realitat hi ha un accentuat creixement dels TEU previ al període de crisi econòmica.

4.3.3. Model del camí aleatori (ARIMA(0,1,0)+c)

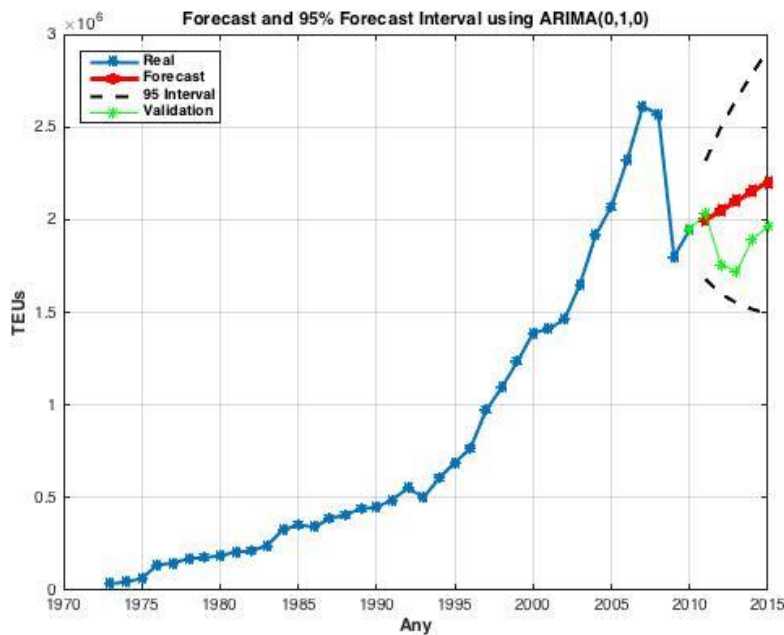


Figura 97 Model del camí aleatori Barcelona verificació 2010

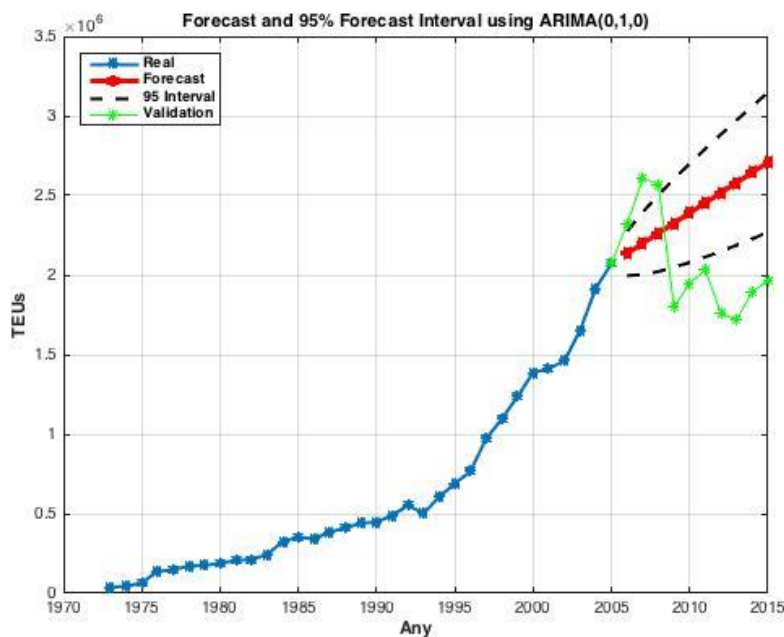


Figura 98 Model del camí aleatori Barcelona verificació 2005

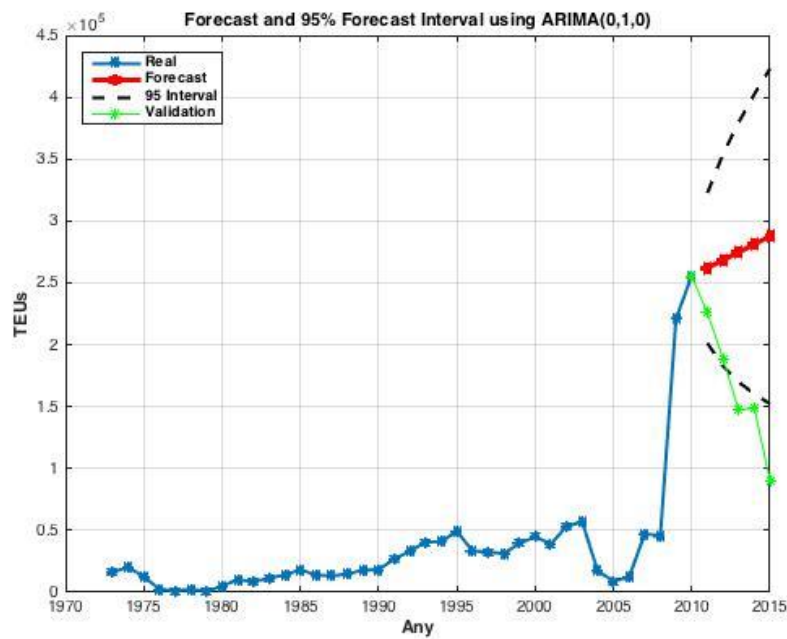


Figura 99 Model del camí aleatori Tarragona verificació 2010

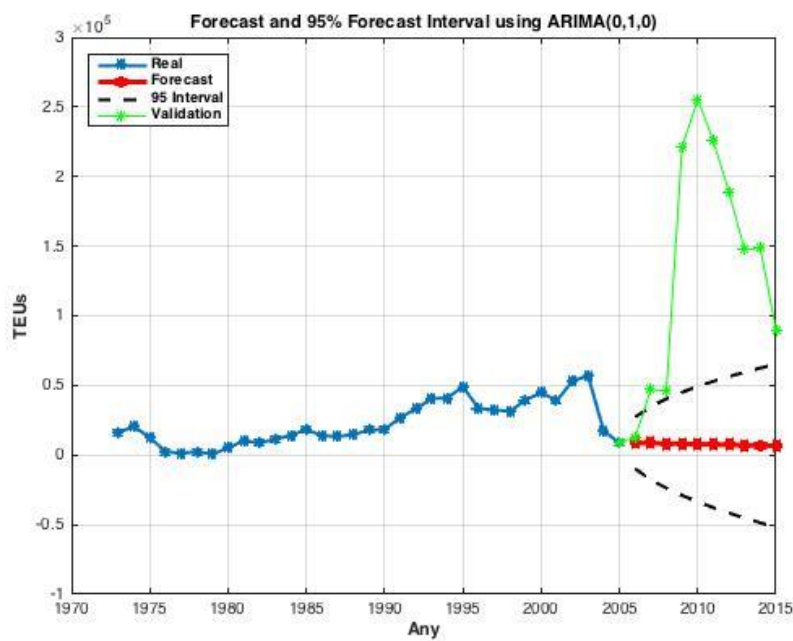


Figura 100 Model del camí aleatori Tarragona verificació 2005

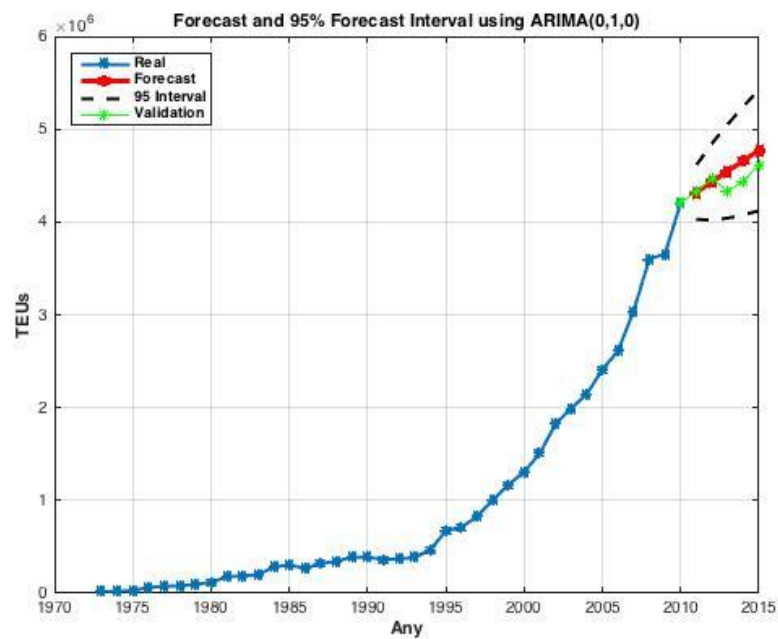


Figura 101 Model del camí aleatori València verificació 2010

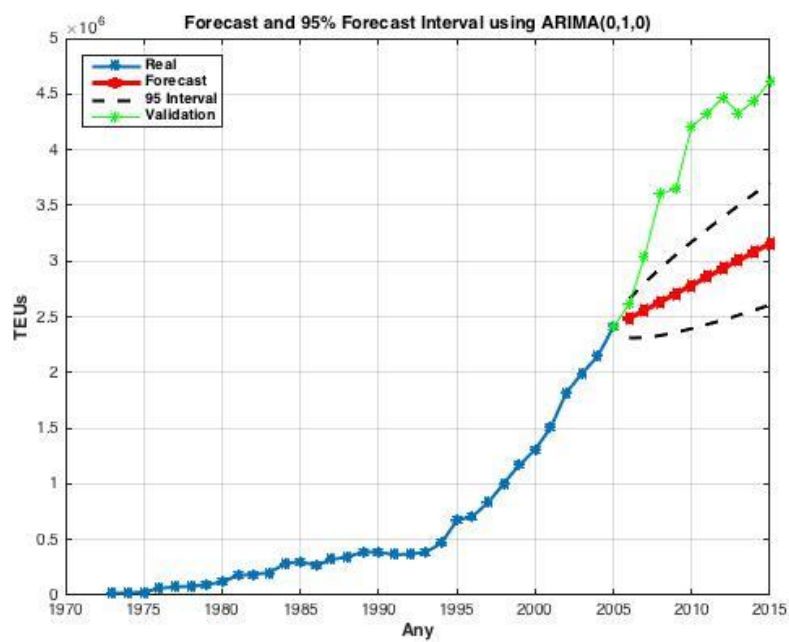


Figura 102 Model del camí aleatori València verificació 2005

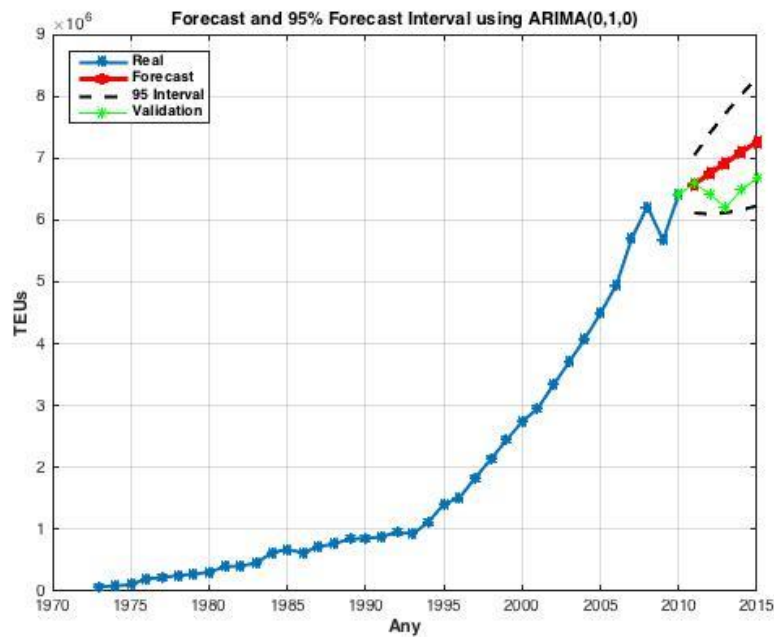


Figura 103 Model del camí aleatori BTV verificació 2010

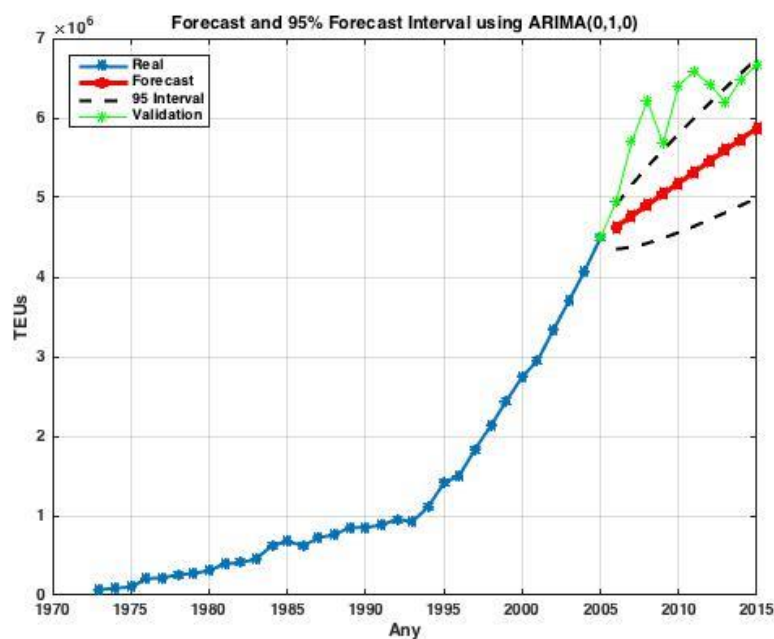


Figura 104 Model del camí aleatori BTV verificació 2005

Aquest és el primer model on els intervals de confiança deixen de ser constants i lineals com ho eren anteriorment. Es pot veure també que en aquest cas quan el període a verificar era a major termini, ha estat molt susceptible l'amplada dels intervals de confiança, ja que sorprenentment en el període del 2005 els intervals eren més estrets que en el cas del 2010 on hi havia menys anys de

predicció. Això pot ser degut a que en els ports els canvis grans en les dades es produeixen a partir del 2005, mentre que abans la tendència era molt lineal en el creixement.

Teòricament s'ha vist que en els models de *random walk* els intervals de confiança no haurien d'estretir-se amb prediccions més a llarg termini, ja que típicament és un model on com ja s'ha explicat es força marcada la regla de l'arrel quadrada del temps, i per tant la tendència general és a que es facin més amplex.

Aquest és un bon exemple del susceptibles que són alguns dels models quan les dades tenen punts d'inflexió o canvis de tendència, modificant el comportament habitual i estimat que aquests haurien de tenir.

Dins d'aquest model es pot destacar la precisió aparent en el cas de València pel període de curt termini, que més endavant es podrà comprovar en l'anàlisi dels errors i que es discutirà en l'apartat següent.

4.3.4. Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)

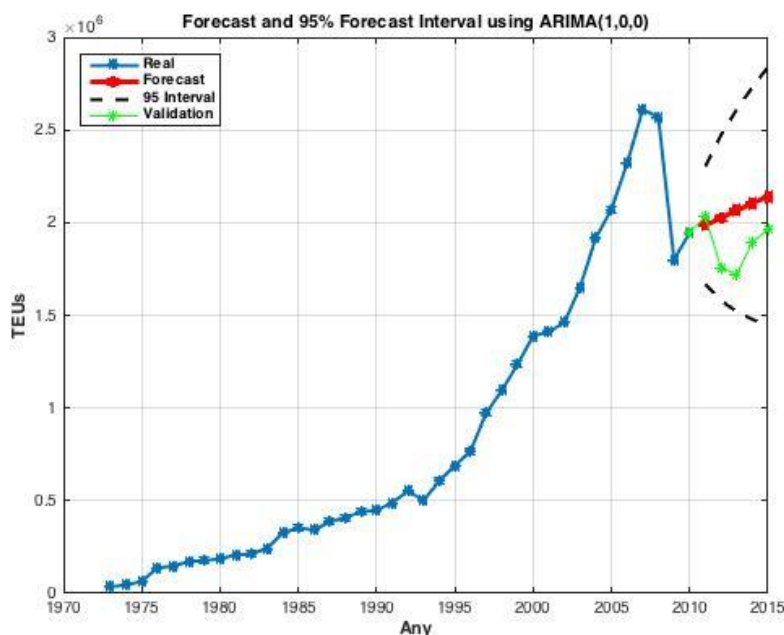


Figura 105 Model AR1 Barcelona verificació 2010

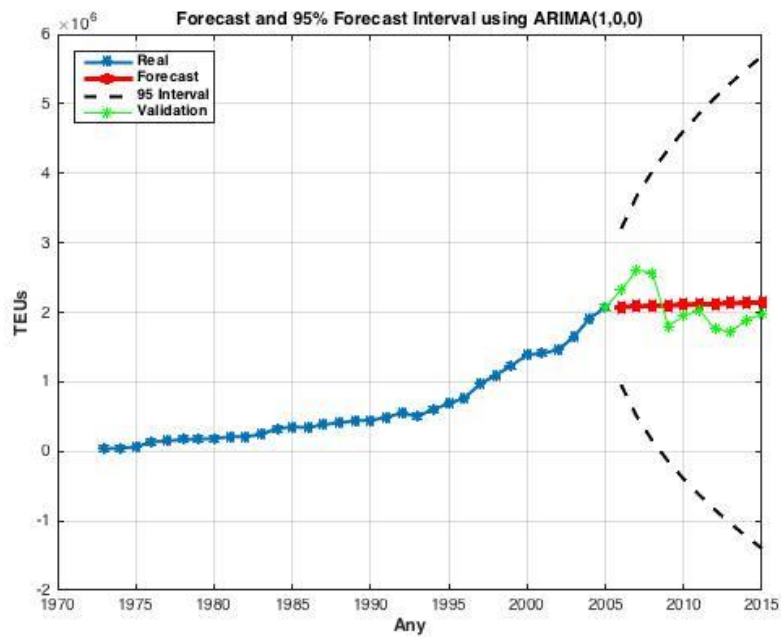


Figura 106 Model AR1 Barcelona verificació 2005

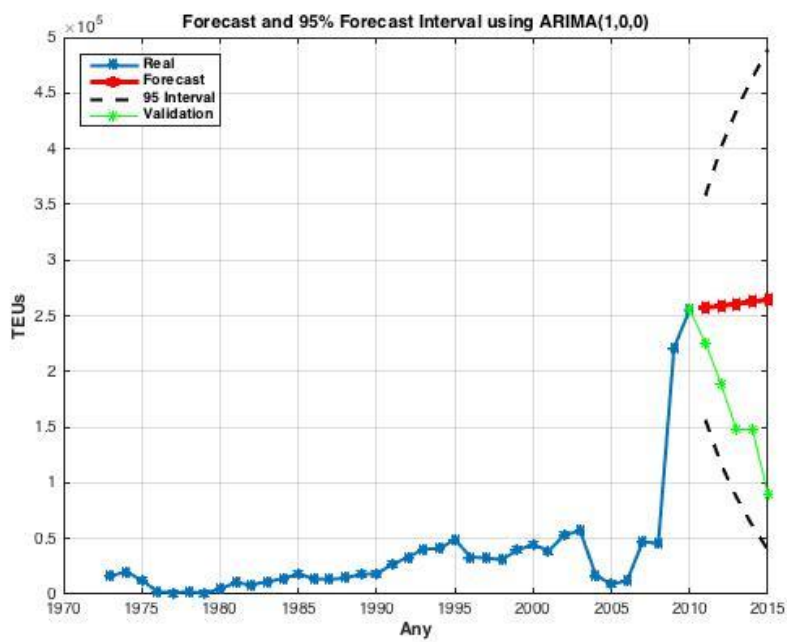


Figura 107 Model AR1 Tarragona verificació 2010

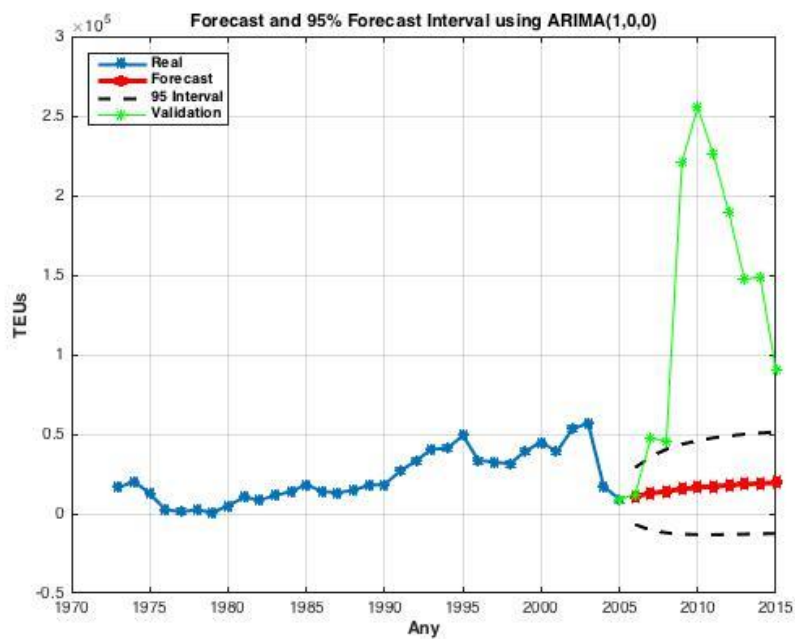


Figura 108 Model AR1 Tarragona verificació 2005

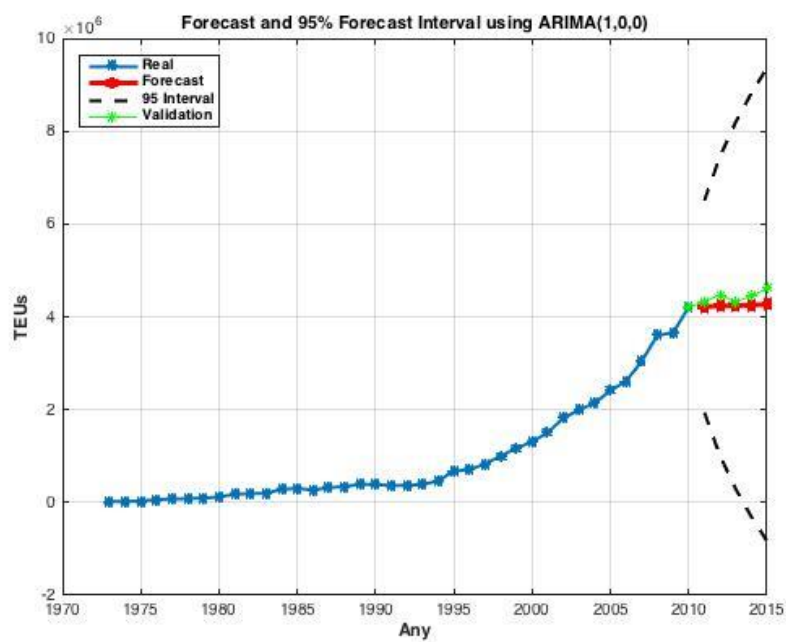


Figura 109 Model AR1 València verificació 2010

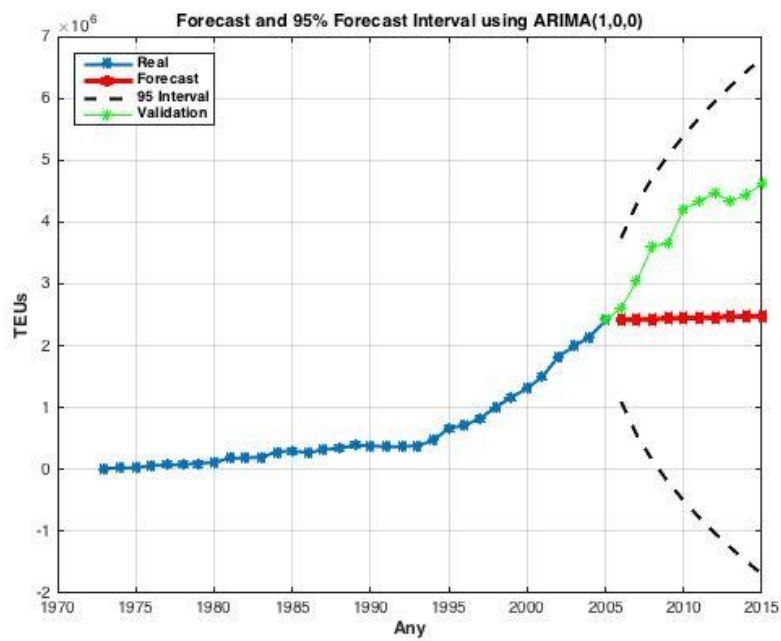


Figura 110 Model AR1 València verificació 2005

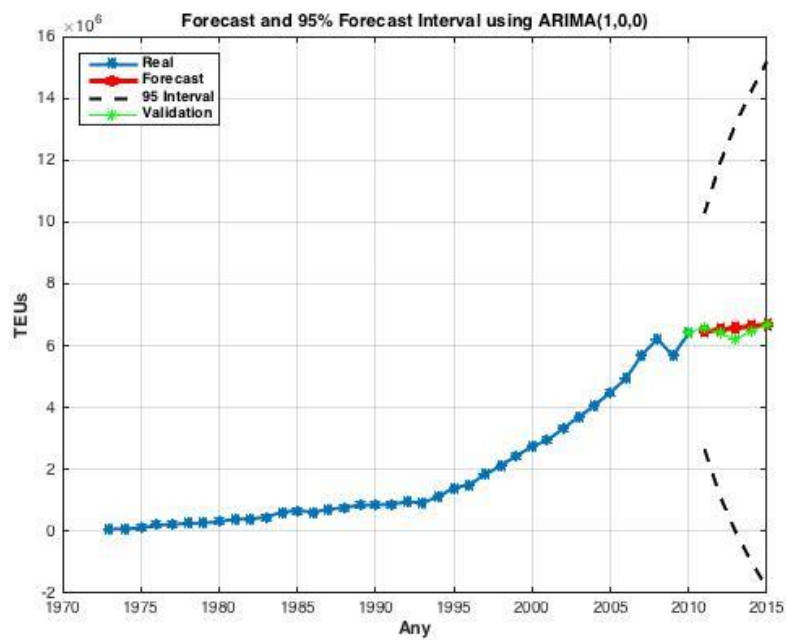


Figura 111 Model AR1 BTV verificació 2010

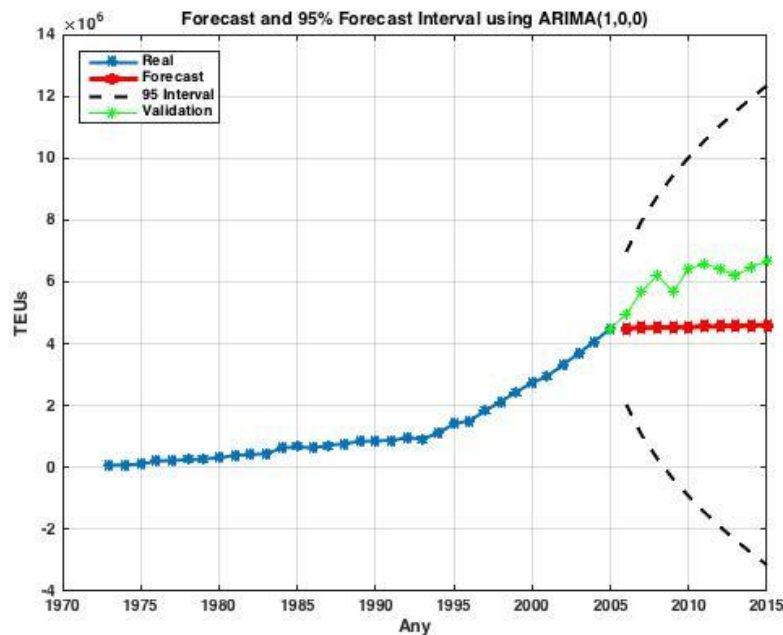


Figura 112 Model AR1 BTV verificació 2005

Per aquests tipus de models autoregressius, s'hi pot veure que han funcionat de forma general més bé que els altres tractats, sobretot amb una millora notòria en el període de verificació més a curt termini del 2010.

En els casos tant de València com l'anàlisi conjunt BTV es poden observar unes prediccions d'alta precisió en la verificació a curt termini com s'ha esmentat, però el model perd consistència al passar a la verificació del 2005. Així mateix, en el cas de Barcelona les prediccions són aparentment acurades en ambdós períodes de verificació.

El cas de Tarragona torna a ser un cas a part. El model torna a respondre de forma errònia als canvis de tendència bruscs que pateix, i és un patró que sembla repetitiu en tots els models tractats fins al moment.

4.3.5. Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)

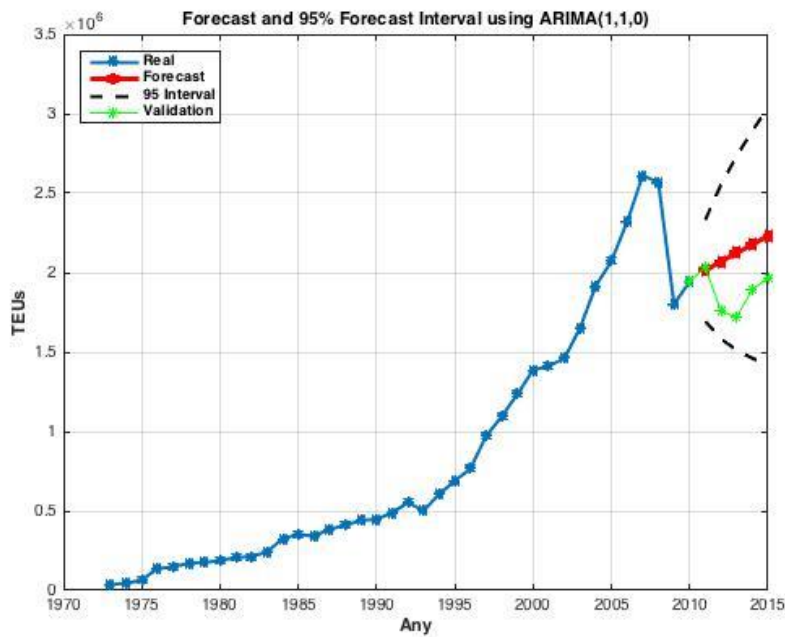


Figura 113 Model AR1 diferenciat Barcelona verificació 2010

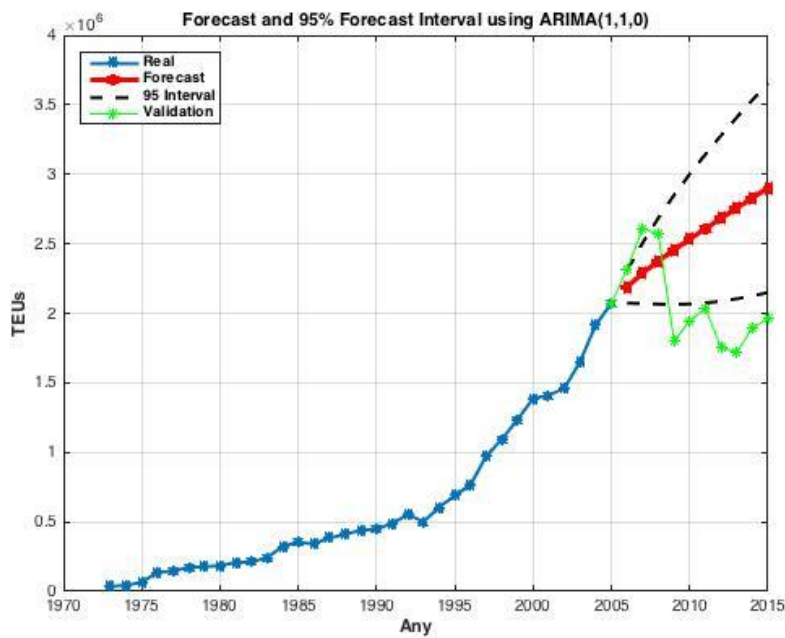


Figura 114 Model AR1 diferenciat Barcelona verificació 2005

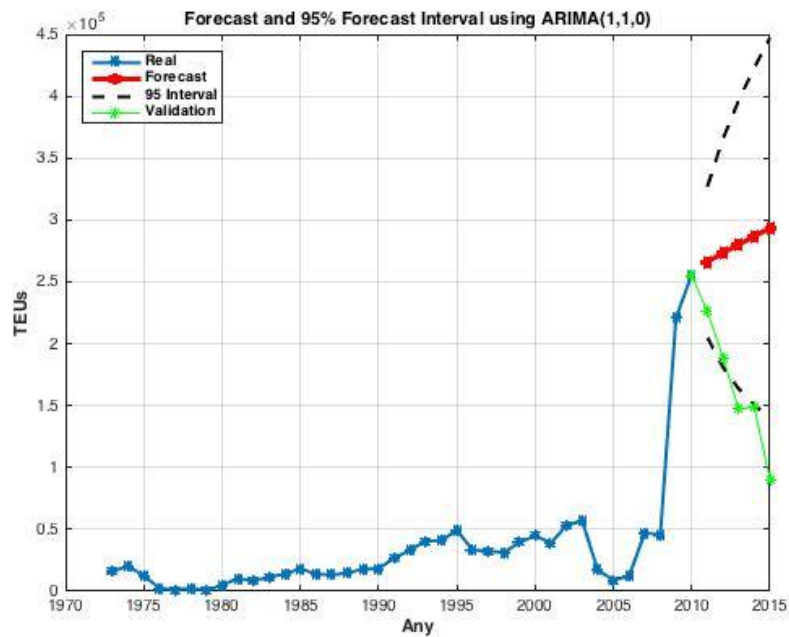


Figura 115 Model AR1 diferenciat Tarragona verificació 2010

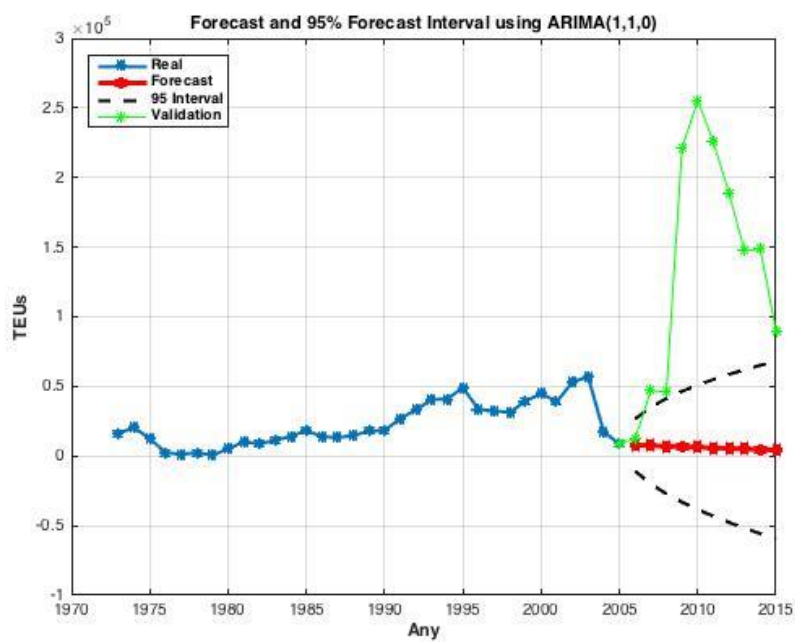


Figura 116 Model AR1 diferenciat Tarragona verificació 2005

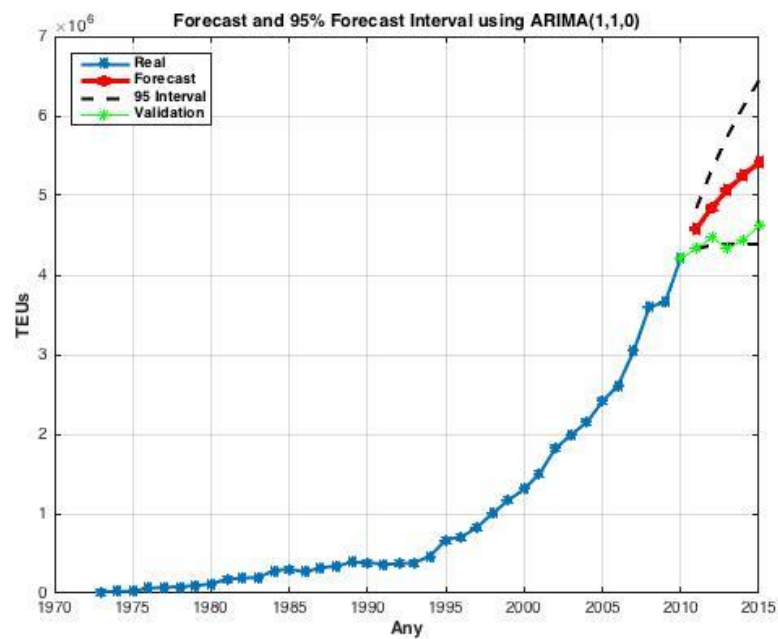


Figura 117 Model AR1 diferenciat València verificació 2010

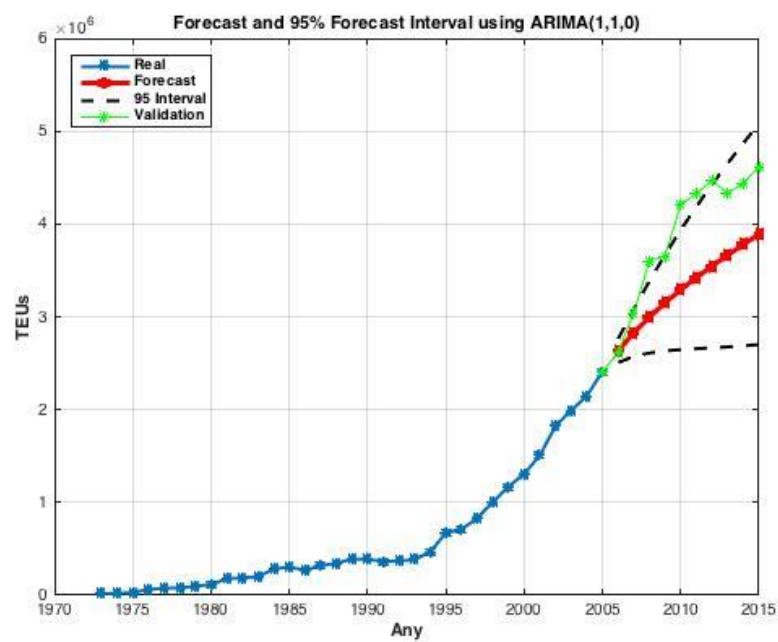


Figura 118 Model AR1 diferenciat València verificació 2005

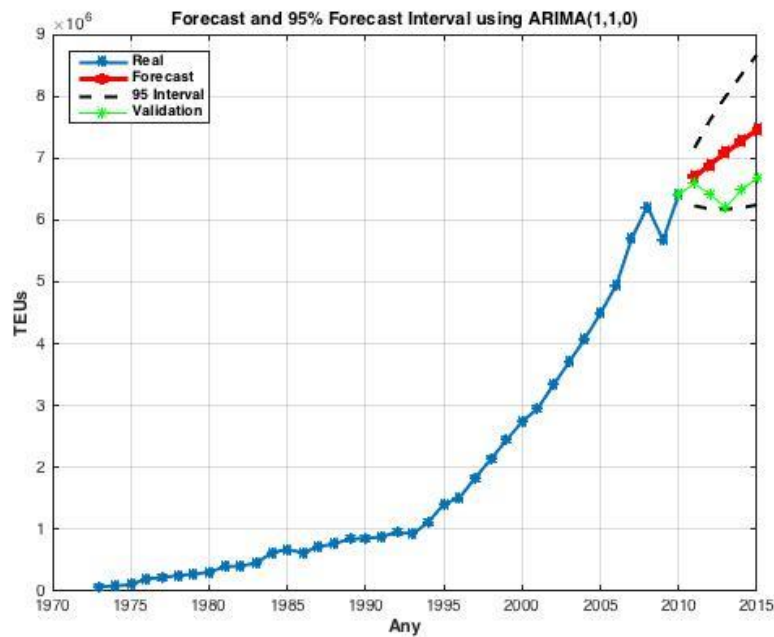


Figura 119 Model AR1 diferenciat BTV verificació 2010

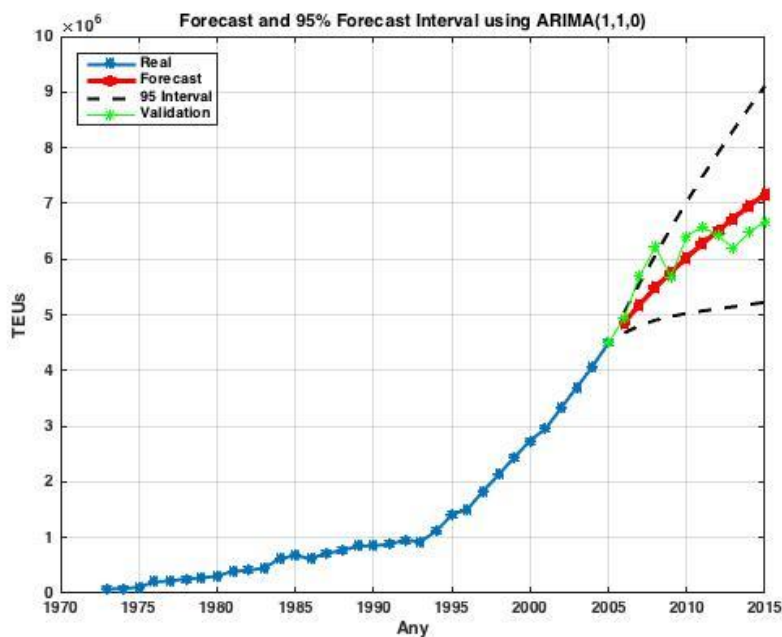


Figura 120 Model AR1 diferenciat BTV verificació 2005

En aquest model observem que no hi ha cap cas puntual en que aparentment el model s'hagi ajustat extraordinàriament bé al curt termini, per tant, sembla haver-se perdut la precisió de la que es disposava per aquest període en el model anterior.

Tenint en compte les prediccions i observant el nombre de TEU tractats, el cas que sembla predir amb major precisió és el cas de l'anàlisi conjunt dels ports BTV, on les verificacions han quedat per dins dels intervals de confiança, seguit del cas de València.

Una vegada més el que més marxa de l'explicació anterior és el port de Tarragona, que sembla no encaixar en cap dels models, almenys fins ara. Podria passar en que cap model dels complexos anés millor que la regressió, que tampoc aporta informació acurada en matèria de prediccions sinó que bàsicament ens orienta de la tendència evolutiva dels TEU.

4.3.6. Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)

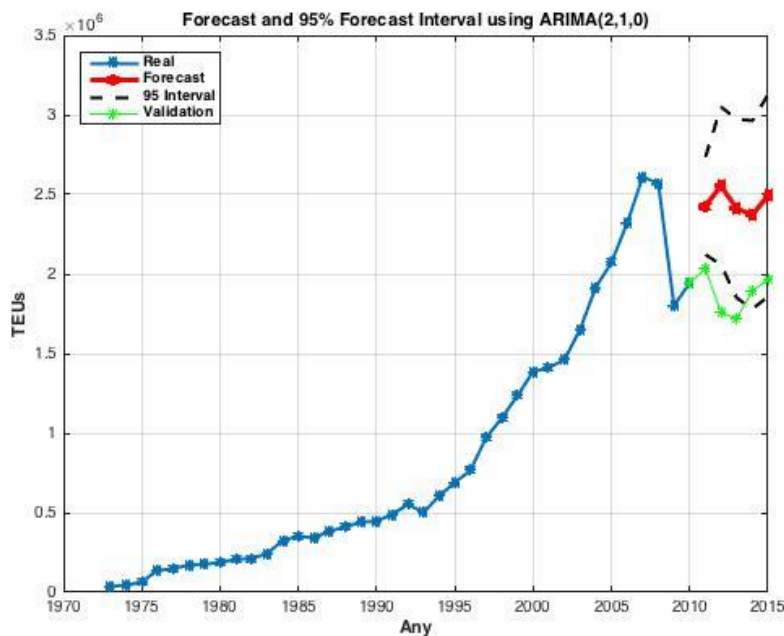


Figura 121 Model AR2 diferenciat Barcelona verificació 2010

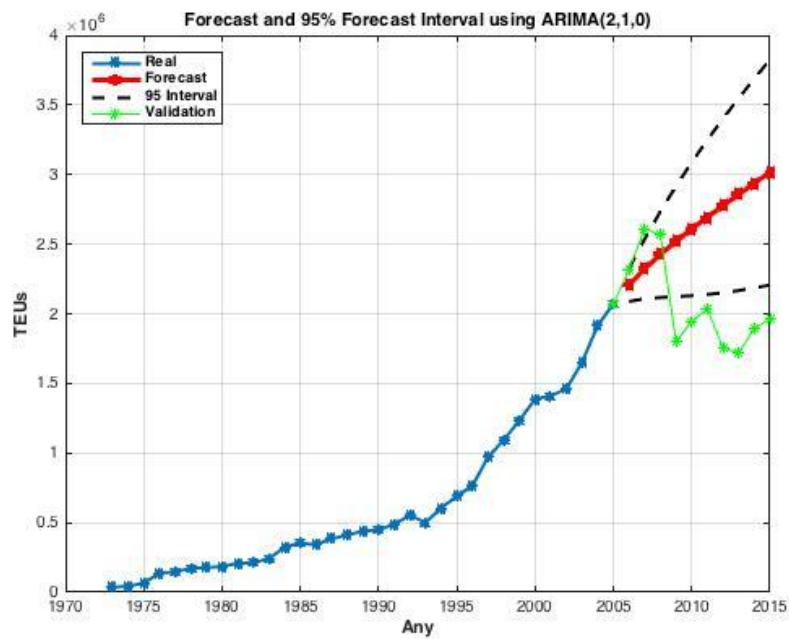


Figura 122 Model AR2 diferenciat Barcelona verificació 2005

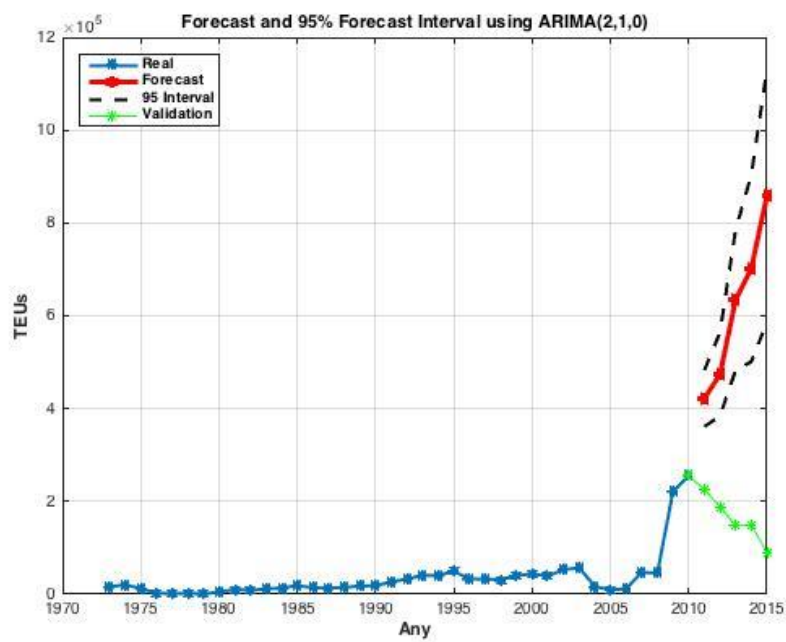


Figura 123 Model AR2 diferenciat Tarragona verificació 2010

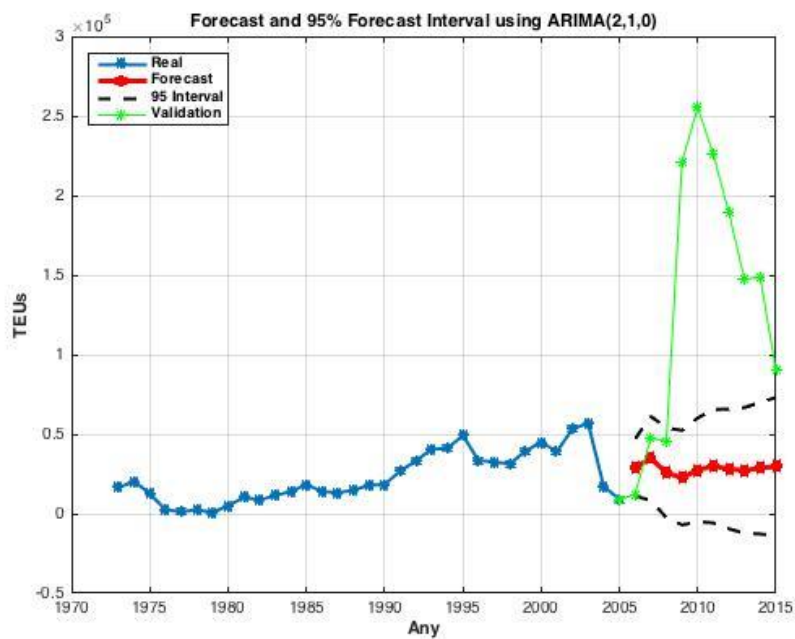


Figura 124 Model AR2 diferenciat Tarragona verificació 2005

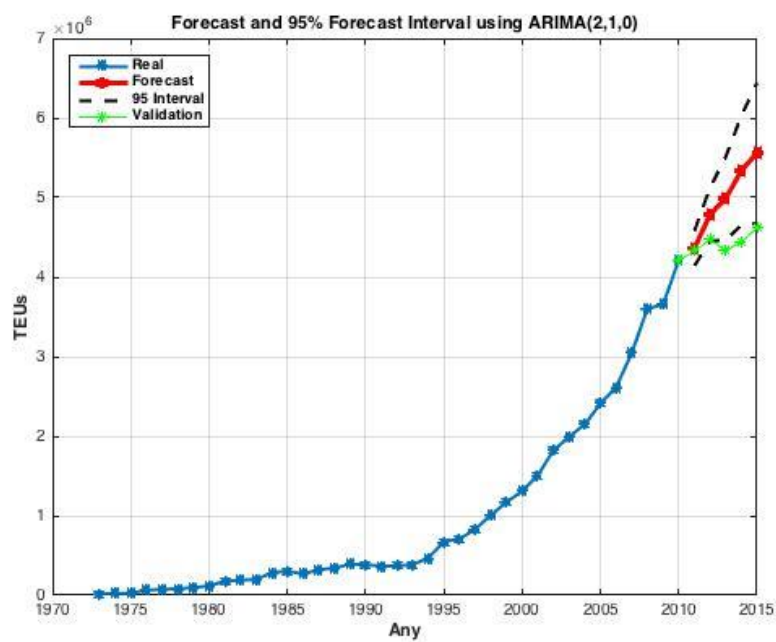


Figura 125 Model AR2 diferenciat València verificació 2010

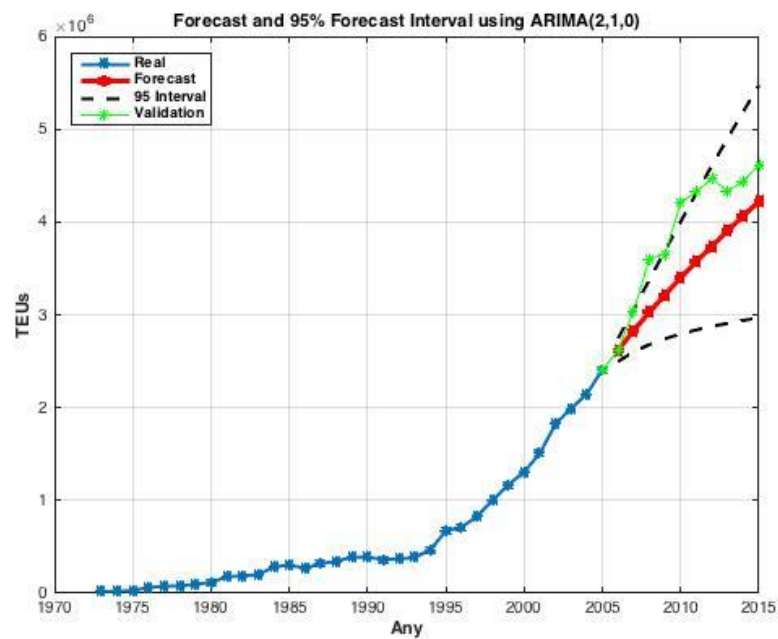


Figura 126 Model AR2 diferenciat València verificació 2005

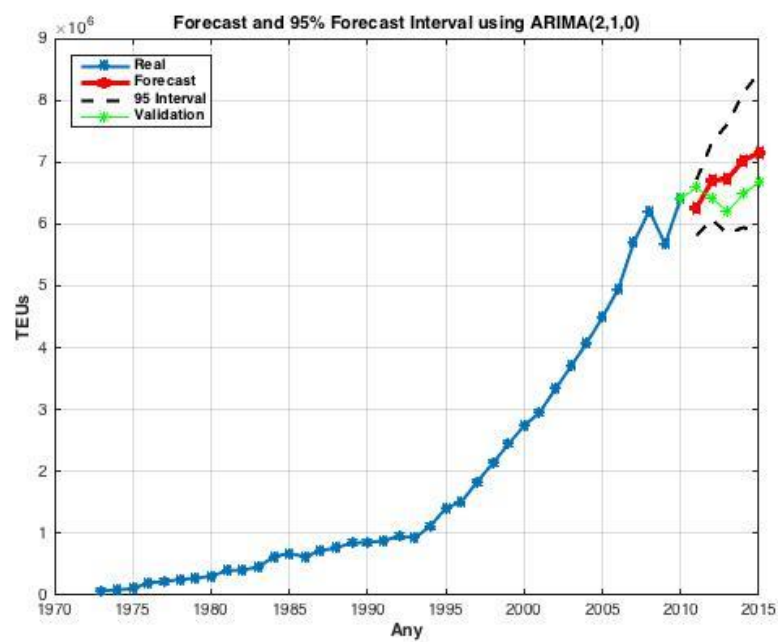


Figura 127 Model AR2 diferenciat BTV verificació 2010

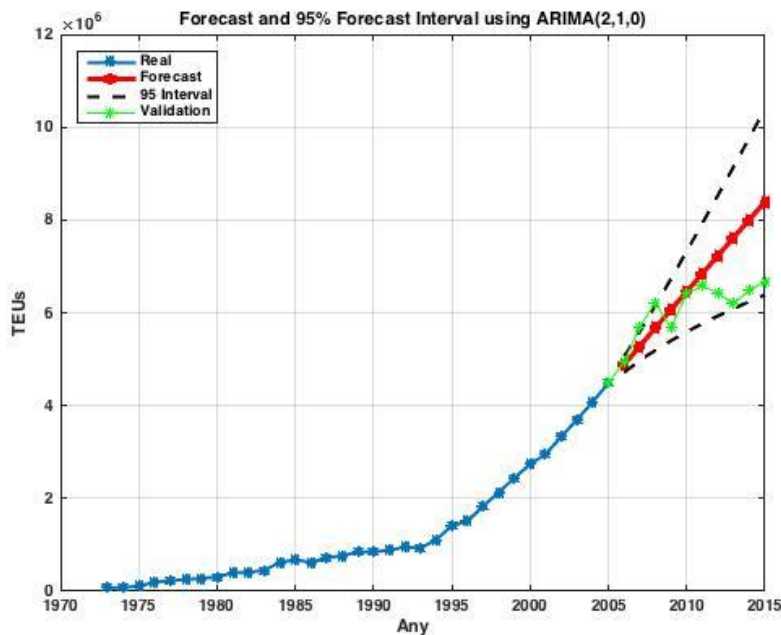


Figura 128 Model AR2 diferenciat BTV verificació 2005

Una tendència clara que es pot observar en aquest model és que els intervals de confiança en general han estat més estrets que en models estudiats anteriorment. Per tant, si ens fos un model vàlid la precisió que aportaria sobre els possibles valors futurs seria gran.

No obstant, en aquest model les prediccions sobretot de Tarragona, però també en Barcelona, han estat més desajustades. Amb Tarragona ja passa de forma habitual però les dades de Barcelona fins ara han permès als diferents models aplicar-se seguint la línia dels altres casos en que els models, amb major o menor precisió, han anat funcionant.

Pel que fa a València funciona de forma acceptable però no millor que en altres models, ja que també es troba molt al límit en les verificacions pel que fa als intervals de confiança mostrats. El cas que si que sembla funcionar bé en les prediccions és en l'anàlisi conjunt dels tres ports, on les prediccions han caigut dins dels límits tot i el poc marge que dóna el model en els intervals per la seva poca amplada.

4.3.7. Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)

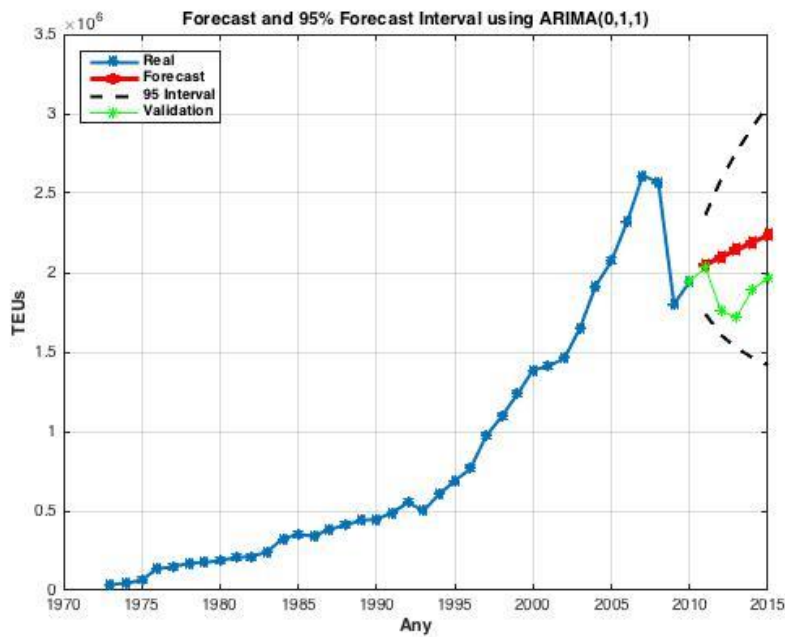


Figura 129 Model d'ajustament exponencial simple Barcelona verificació 2010

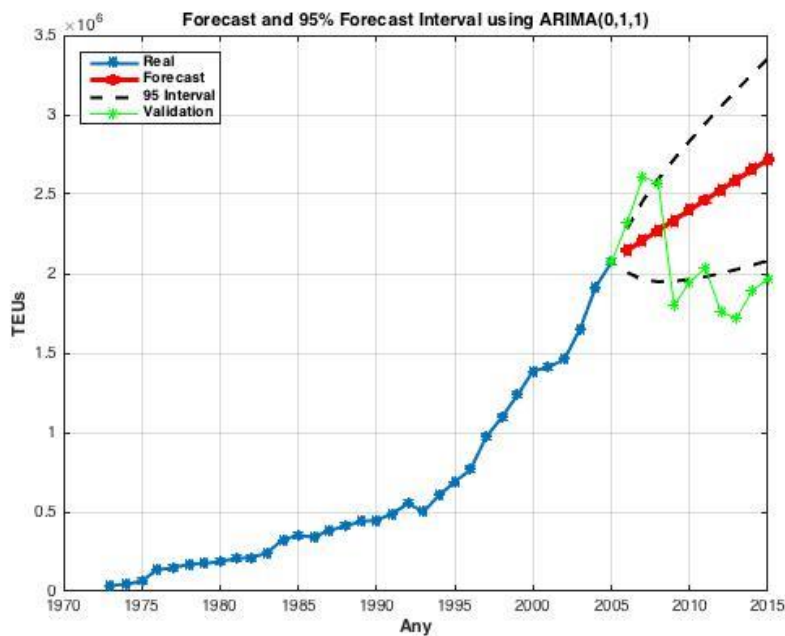


Figura 130 Model d'ajustament exponencial simple Barcelona verificació 2005

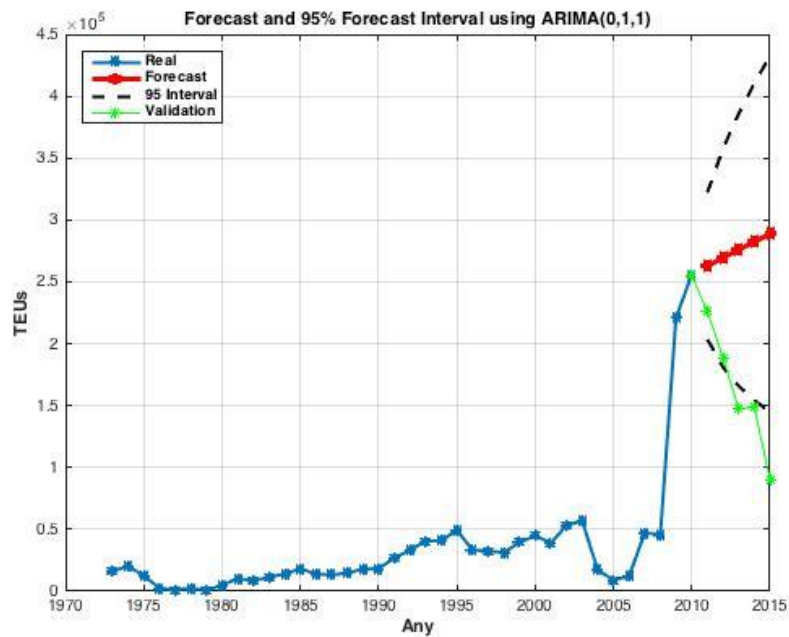


Figura 131 Model d'ajustament exponencial simple Tarragona verificació 2010

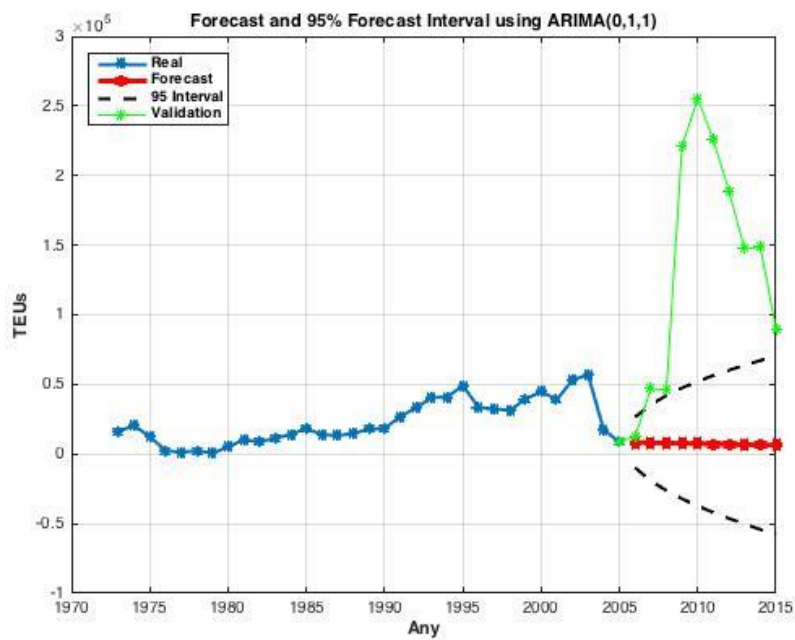


Figura 132 Model d'ajustament exponencial simple Tarragona verificació 2005

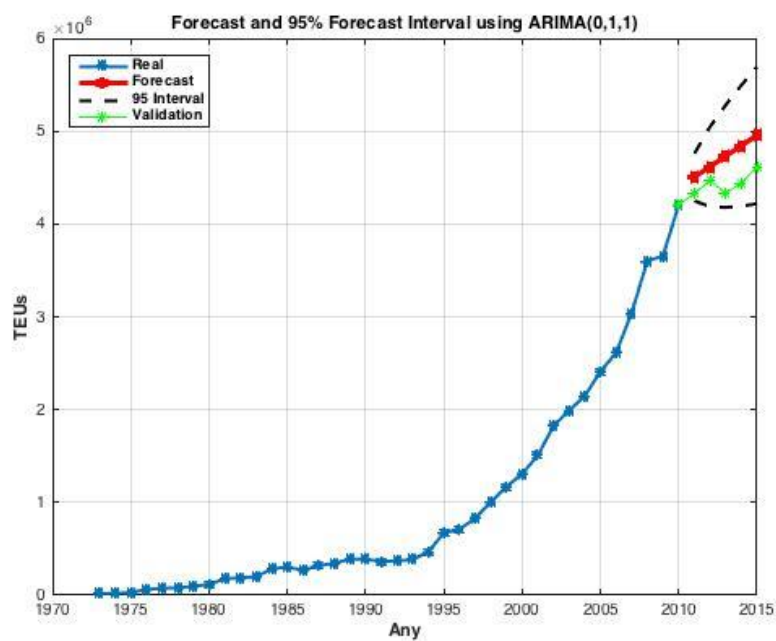


Figura 133 Model d'ajustament exponencial simple València verificació 2010

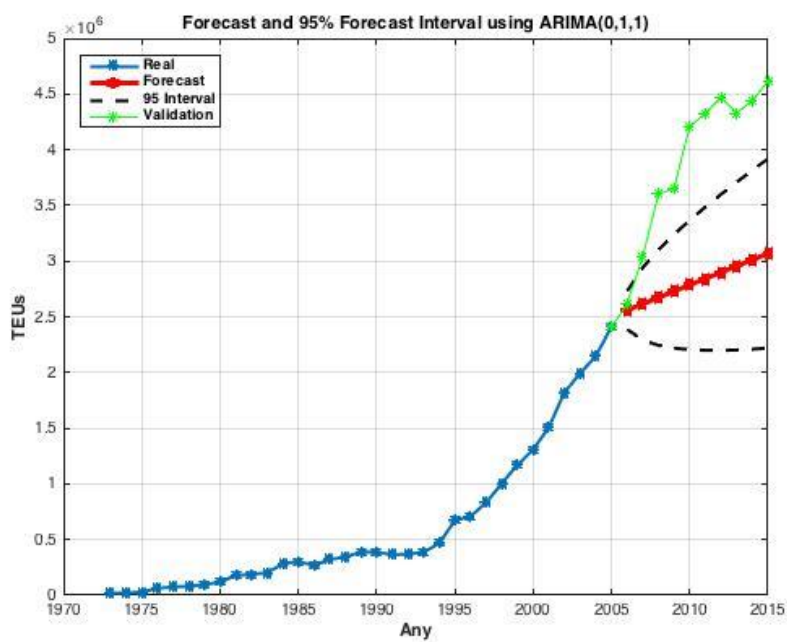


Figura 134 Model d'ajustament exponencial simple València verificació 2005

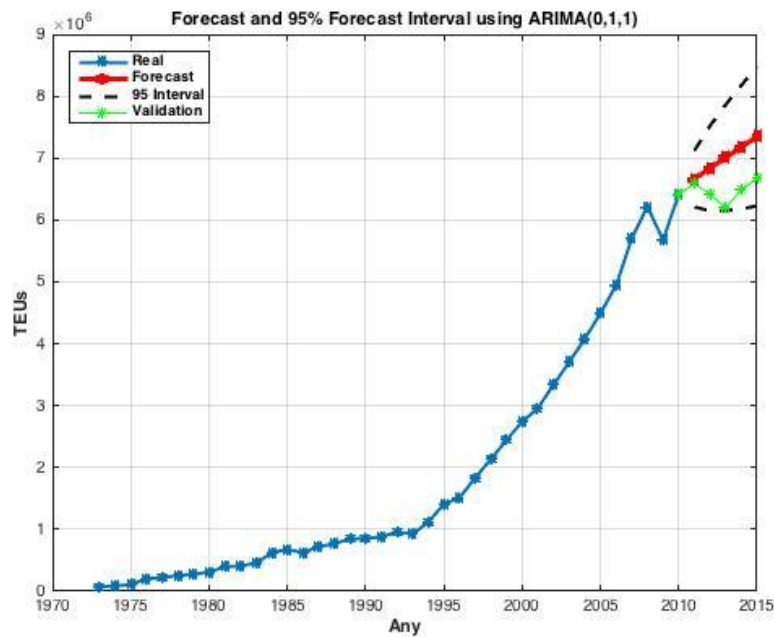


Figura 135 Model d'ajustament exponencial simple BTV verificació 2010

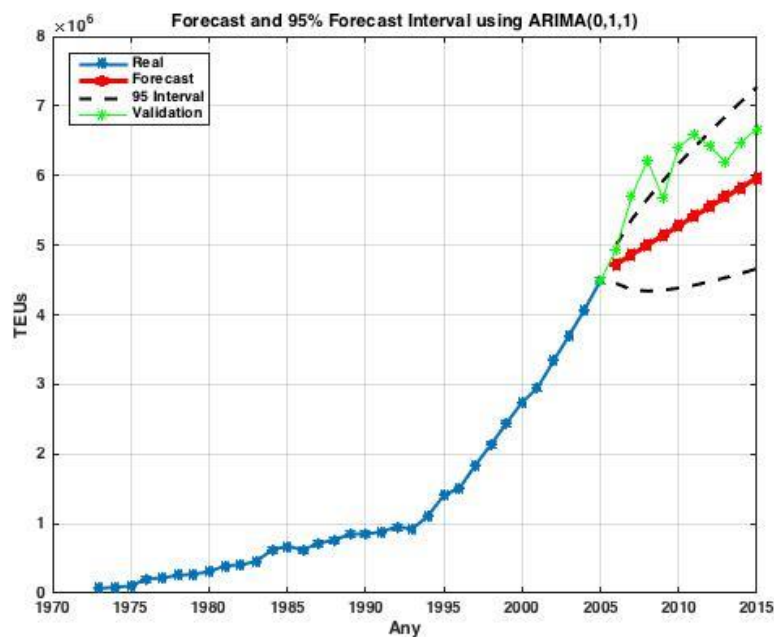


Figura 136 Model d'ajustament exponencial simple BTV verificació 2005

Es pot observar que les prediccions en els diferents casos són acceptables i comparables a altres models pel que fa al període de verificació de 2010, és a dir, curt termini, però que tendeix a fallar en quasi tots en el període de validació de 2005, a llarg termini.

L'únic cas en que les prediccions es mantenen més o menys parelles en els dos períodes de verificació és en l'anàlisi conjunt BTV, i el cas en que falla en ambdós torna a ser Tarragona.

Sorprèn que en molts dels casos que no han acabat de funcionar en els diferents ports, han funcionat millor en el cas dels ports conjunts, ja que tal i com s'ha explicat en l'apartat anterior de descripció de dades, el model que menys variacions tenia era el de València.

4.3.8. Taules resum dels errors

RESULTATS MODELS BARCELONA	Període	RMSE	MAE	MAPE
Regressió lineal	Validació 2010	359 900.00	317 720.00	17.33
	Validació 2005	506 060.00	380 550.00	16.85
Model de la mitjana (ARIMA(0,0,0)+c)	Validació 2010	1 028 600.00	1 021 600.00	54.34
	Validació 2005	1 453 900.00	1 420 700.00	68.28
Model de camí aleatori (ARIMA(0,1,0)+c)	Validació 2010	266 630.00	241 310.00	13.27
	Validació 2005	581 350.00	540 480.00	27.90
Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	Validació 2010	232 140.00	209 930.00	11.57
	Validació 2005	327 700.00	297 960.00	14.45
Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)	Validació 2010	286 280.00	256 580.00	14.12
	Validació 2005	703 560.00	629 790.00	33.04
Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)	Validació 2010	595 910.00	577 640.00	31.41
	Validació 2005	776 950.00	683 660.00	36.07
Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	Validació 2010	301 870.00	269 270.00	14.83
	Validació 2005	587 520.00	544 430.00	28.15

Taula 1 Taula d'errors obtinguda dels exercicis de verificació a les dades del port de Barcelona

RESULTATS MODELS TARRAGONA	Període	RMSE	MAE	MAPE
Regressió lineal	Validació 2010	104 390.00	88 253.00	44.10
	Validació 2005	115 260.00	90 368.00	100.58
Model de la mitjana (ARIMA(0,0,0)+c)	Validació 2010	133 060.00	125 030.00	75.91
	Validació 2005	141 240.00	117 600.00	79.12
Model de camí aleatori (ARIMA(0,1,0)+c)	Validació 2010	126 980.00	114 720.00	90.82
	Validació 2005	153 890.00	130 510.00	86.13
Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	Validació 2010	111 600.00	100 740.00	79.83
	Validació 2005	145 820.00	122 010.00	77.38
Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)	Validació 2010	131 860.00	119 860.00	94.46
	Validació 2005	155 360.00	132 120.00	88.11
Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)	Validació 2010	500 820.00	458 090.00	359.26
	Validació 2005	137 410.00	113 240.00	78.71
Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	Validació 2010	127 900.00	115 740.00	91.52
	Validació 2005	154 210.00	130 890.00	86.76

Taula 2 Taula d'errors obtinguda dels exercicis de verificació a les dades del port de Tarragona

RESULTATS MODELS VALÈNCIA	Període	RMSE	MAE	MAPE
Regressió lineal	Validació 2010	1 488 200.00	1 485 900.00	33.81
	Validació 2005	1 969 300.00	1 883 800.00	48.36
Model de la mitjana (ARIMA(0,0,0)+c)	Validació 2010	3 447 400.00	3 445 800.00	77.66
	Validació 2005	3 369 700.00	3 308 000.00	83.65
Model de camí aleatori (ARIMA(0,1,0)+c)	Validació 2010	155 860.00	127 550.00	2.87
	Validació 2005	1 197 800.00	1 109 200.00	26.84
Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	Validació 2010	216 380.00	194 310.00	4.33
	Validació 2005	1 607 400.00	1 481 500.00	35.74
Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)	Validació 2010	642 110.00	598 490.00	13.45
	Validació 2005	677 320.00	614 490.00	14.79
Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)	Validació 2010	666 770.00	568 390.00	12.71
	Validació 2005	528 630.00	471 550.00	11.49
Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	Validació 2010	313 270.00	293 200.00	6.61
	Validació 2005	1 221 500.00	1 116 200.00	26.79

Taula 3 Taula d'errors obtinguda dels exercicis de verificació a les dades del port de València

RESULTATS MODELS BTV	Període	RMSE	MAE	MAPE
Regressió lineal	Validació 2010	1 283 800.00	1 254 200.00	19.39
	Validació 2005	2 279 000.00	2 240 800.00	37.21
Model de la mitjana (ARIMA(0,0,0)+c)	Validació 2010	4 595 300.00	4 592 500.00	70.96
	Validació 2005	4 870 900.00	4 844 200.00	78.87
Model de camí aleatori (ARIMA(0,1,0)+c)	Validació 2010	522 420.00	454 550.00	7.07
	Validació 2005	932 060.00	879 610.00	14.16
Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	Validació 2010	194 170.00	151 590.00	2.39
	Validació 2005	1 651 600.00	1 578 800.00	25.22
Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)	Validació 2010	674 180.00	610 130.00	9.48
	Validació 2005	426 890.00	371 350.00	5.96
Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)	Validació 2010	446 460.00	433 100.00	6.71
	Validació 2005	931 150.00	720 780.00	11.39
Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	Validació 2010	596 780.00	535 900.00	8.33
	Validació 2005	841 180.00	782 660.00	12.56

Taula 4 Taula d'errors obtinguda dels exercicis de verificació a les dades conjuntes dels ports BTV

5. DISCUSSIÓ

Presentats els diferents resultats obtinguts en els diferents casos i models, es pot fer una valoració general respecte a quin o quins models han funcionat millor i si hi poden tenir una cabuda real pel que fa al *forecasting* de tràfic portuari.

Els arguments que s'aniran desenvolupant tindran en compte tot el que s'ha vist fins ara en la memòria, des del punt de vista teòric, resultats, períodes de verificació analitzats, residuals, autocorrelacions, estadístics i errors dels diferents models tractats. Sempre sense perdre l'objectiu de trobar un model de predicció que funcioni en els nostres casos.

Per tal de fer més entenedores les taules amb les que es procedirà a l'anàlisi dels errors, aquestes es tornaran a presentar amb uns valors remarcats en verd. Aquests valors es correspondran als valors més baixos en cada cas, és a dir, els valors d'error mínim entre models. D'aquesta manera s'obtindrà el valor més baix per a cada error i cada període, ja que es marcaran tant pel període de curt termini a partir del 2010, com pel període a llarg termini a partir del 2005.

Aquestes taules seran útils en referències posteriors:

RESULTATS MODELS BARCELONA	Període	RMSE	MAE	MAPE
Regressió lineal	Validació 2010	359 900.00	317 720.00	17.33
	Validació 2005	506 060.00	380 550.00	16.85
Model de la mitjana (ARIMA(0,0,0)+c)	Validació 2010	1 028 600.00	1 021 600.00	54.34
	Validació 2005	1 453 900.00	1 420 700.00	68.28
Model de camí aleatori (ARIMA(0,1,0)+c)	Validació 2010	266 630.00	241 310.00	13.27
	Validació 2005	581 350.00	540 480.00	27.90
Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	Validació 2010	232 140.00	209 930.00	11.57
	Validació 2005	327 700.00	297 960.00	14.45
Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)	Validació 2010	286 280.00	256 580.00	14.12
	Validació 2005	703 560.00	629 790.00	33.04
Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)	Validació 2010	595 910.00	577 640.00	31.41
	Validació 2005	776 950.00	683 660.00	36.07
Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	Validació 2010	301 870.00	269 270.00	14.83
	Validació 2005	587 520.00	544 430.00	28.15

Taula 5 Taula tipificada per a l'anàlisi d'errors en les verificacions de les dades del port de Barcelona

RESULTATS MODELS TARRAGONA	Període	RMSE	MAE	MAPE
Regressió lineal	Validació 2010	104 390.00	88 253.00	44.10
	Validació 2005	115 260.00	90 368.00	100.58
Model de la mitjana (ARIMA(0,0,0)+c)	Validació 2010	133 060.00	125 030.00	75.91
	Validació 2005	141 240.00	117 600.00	79.12
Model de camí aleatori (ARIMA(0,1,0)+c)	Validació 2010	126 980.00	114 720.00	90.82
	Validació 2005	153 890.00	130 510.00	86.13
Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	Validació 2010	111 600.00	100 740.00	79.83
	Validació 2005	145 820.00	122 010.00	77.38
Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)	Validació 2010	131 860.00	119 860.00	94.46
	Validació 2005	155 360.00	132 120.00	88.11
Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)	Validació 2010	500 820.00	458 090.00	359.26
	Validació 2005	137 410.00	113 240.00	78.71
Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	Validació 2010	127 900.00	115 740.00	91.52
	Validació 2005	154 210.00	130 890.00	86.76

Taula 6 Taula tipificada per a l'anàlisi d'errors en les verificacions de les dades del port de Tarragona

RESULTATS MODELS VALÈNCIA	Període	RMSE	MAE	MAPE
Regressió lineal	Validació 2010	1 488 200.00	1 485 900.00	33.81
	Validació 2005	1 969 300.00	1 883 800.00	48.36
Model de la mitjana (ARIMA(0,0,0)+c)	Validació 2010	3 447 400.00	3 445 800.00	77.66
	Validació 2005	3 369 700.00	3 308 000.00	83.65
Model de camí aleatori (ARIMA(0,1,0)+c)	Validació 2010	155 860.00	127 550.00	2.87
	Validació 2005	1 197 800.00	1 109 200.00	26.84
Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	Validació 2010	216 380.00	194 310.00	4.33
	Validació 2005	1 607 400.00	1 481 500.00	35.74
Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)	Validació 2010	642 110.00	598 490.00	13.45
	Validació 2005	677 320.00	614 490.00	14.79
Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)	Validació 2010	666 770.00	568 390.00	12.71
	Validació 2005	528 630.00	471 550.00	11.49
Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	Validació 2010	313 270.00	293 200.00	6.61
	Validació 2005	1 221 500.00	1 116 200.00	26.79

Taula 7 Taula tipificada per a l'anàlisi d'errors en les verificacions de les dades del port de València

RESULTATS MODELS BTV	Període	RMSE	MAE	MAPE
Regressió lineal	Validació 2010	1 283 800.00	1 254 200.00	19.39
	Validació 2005	2 279 000.00	2 240 800.00	37.21
Model de la mitjana (ARIMA(0,0,0)+c)	Validació 2010	4 595 300.00	4 592 500.00	70.96
	Validació 2005	4 870 900.00	4 844 200.00	78.87
Model de camí aleatori (ARIMA(0,1,0)+c)	Validació 2010	522 420.00	454 550.00	7.07
	Validació 2005	932 060.00	879 610.00	14.16
Model autoregressiu de primer ordre (ARIMA(1,0,0)+c)	Validació 2010	194 170.00	151 590.00	2.39
	Validació 2005	1 651 600.00	1 578 800.00	25.22
Model autoregressiu de primer ordre diferenciat (ARIMA(1,1,0)+c)	Validació 2010	674 180.00	610 130.00	9.48
	Validació 2005	426 890.00	371 350.00	5.96
Model autoregressiu de segon ordre diferenciat (ARIMA(2,1,0)+c)	Validació 2010	446 460.00	433 100.00	6.71
	Validació 2005	931 150.00	720 780.00	11.39
Model d'ajustament exponencial simple amb constant (ARIMA(0,1,1)+c)	Validació 2010	596 780.00	535 900.00	8.33
	Validació 2005	841 180.00	782 660.00	12.56

Taula 8 Taula tipificada per a l'anàlisi d'errors en les verificacions de les dades conjuntes dels ports BTV

Seguint l'ordre en que s'ha presentat els resultats al llarg d'aquesta memòria, el primer model a discutir serà el de regressió lineal o de regressió simple. Tal i com s'hi ha referit al llarg del treball, aquest ha estat el model més simple que hem utilitzat, i l'objectiu és comparar-lo amb els altres models ARIMA complexos per observar si la complexitat afegida compensa en matèria de qualitat de precisió.

La solució més lògica seria trobar-nos en que ha estat un model amb pitjor ajust en relació amb els altres, però fixant-nos en els altres es pot veure que de forma sorprenent que en el cas de Tarragona en les verificacions de pràcticament ambdós períodes ha estat les que han aportat major precisió. Si bé això és cert, és obligat també fer referència que les precisions obtingudes en el cas de Tarragona han estat molt fora de lloc i que el valor que s'hi obté fa que sigui un model clarament refutable del qual no ens podríem fiar a l'hora de fer prediccions.

El fet que hagi sigut el de menor error en Tarragona ve de la pròpia naturalesa de les dades, i és que com s'ha referit anteriorment en l'apartat de descripció de dades, és el port amb més variació de TEU durant els períodes, previ, durant, i posterior a la crisi, fent-se presents molts canvis de tendència amb punts forts d'inflexió que han dificultat molt les prediccions amb els models univariants que s'han tractat en aquesta memòria.

També és interessant puntualitzar que els valors més baixos en matèria d'errors de la regressió s'han obtingut en el cas de Barcelona, amb uns errors propers a altres models que han estat considerats com acceptables. Per tant, per aquest

port la labor del model com a merament explicatiu pot ser considerada com a vàlida, sense perdre de vista que altres models hi han funcionat millor.

El següent model ha estat el model de la mitjana, tractant-se del model de construcció més simple dins del grup dels ARIMA. Si bé és cert que no ha aportat qualitat de predicció en les dades dels diferents ports, sí que és un model que ha servit molt bé per entendre la naturalesa de les dades que s'han tractat.

En l'anàlisi del model de la mitjana hem observat clarament les traces de tendència autoregressiva que s'han manifestat al llarg de tots els models. S'ha vist per la caiguda lenta en el gràfic d'autocorrelació de la mostra i el canvi brusc en els primers *lags* del gràfic d'autocorrelacions parcials de la mostra, amb marcades autocorrelacions positives. També s'hi ha observat un gràfic de residuals que no mostrava l'aparença de variables aleatòries idènticament distribuïdes, per tant, de forma teòrica aquest model seria estadísticament ajustable. Per recordar aquestes apreciacions val amb tenir present [l'Il·lustració 5](#) presentada en l'explicació dels models ARIMA.

El següent model analitzat ha estat el de camí aleatori, que en ha permès corregir el desajust en el gràfic de residuals gràcies al grau de diferenciació aportat, tal i com hem vist en l'apartat de resultats. En aquest model ja es pot observar una gran reducció pel que fa als errors en la majoria dels ports tractats, deixant de banda el fet que Tarragona com hem analitzat ha seguit sempre un patronatge diferent.

S'hi veu que els resultats obtinguts al cas de Barcelona no han estat desmesurats però no ha estat el millor model, però tal i com apuntàvem en els períodes de verificació es confirma en l'anàlisi dels errors que el cas de València pel període a curt termini ha estat realment el més acurat per aquest port. El punt negatiu ha estat que tot i ser el més precís a curt termini, el model no ha resultat robust quan hem ampliat el període de verificació. Això porta a pensar-lo com a un model vàlid per aquest port, però amb cautela de no caure en el parany d'ampliar el termini de la nostra predicció ja que el torna un model de poca utilitat.

Pel cas dels ports conjunts el model de camí aleatori ha obtingut uns errors força acceptables i amb una pèrdua de precisió a l'augmentar el termini dins del normal, però no ha estat el millor model per aquest cas.

El següent model és l'autoregressiu de primer ordre, que com s'ha vist en anteriors apartats es mostra estadísticament acceptable en els seus valors i ha tingut unes precisions força acceptables que ara es compararan en l'anàlisi dels errors. Els raonaments que s'han anat desenvolupant en el treball porten a pensar que un dels models autoregressius que s'han tractat hauria de ser el de major ajust per la naturalesa de les diferents dades.

És el model que millor ha funcionat en el cas de Barcelona, tant en el període a curt termini com en el llarg, mostrant-se consistent amb el temps i amb una pèrdua de precisió a l'augmentar el període que no ha estat notòria. Cal apuntar que no s'ha obtingut un valor de fiabilitat inferior al deu per cent pel que fa al MAPE però tot i així per ser els valors més baixos s'ha de considerar aquest model com el vàlid.

En els casos de València i els ports conjunts aquest model ha aportat una precisió a curt termini molt gran, essent el més precís a curt termini per l'anàlisi conjunt BTV. No obstant, torna a trobar-se la poca consistència a l'augmentar el termini de les prediccions. Aquest augment no és preocupant ni és un motiu per descartar el model, ja que és una particularitat que ja s'anticipava en relació a l'anàlisi de sèries temporals, però és cert que en el cas de poder trobar un model que aportí informació precisa tant a curt com a llarg termini haurà de ser considerat com a vàlid o com a ser analitzat en relació a altres com aquest que aportin gran precisió en algun dels períodes.

Pel cas del model autoregressiu de primer ordre diferenciat, s'ha observat en l'apartat de resultats que no semblava aportar cap millora estadística clara i que cap de les prediccions semblava ser prou acurada com per ser digna de destacar.

En l'anàlisi dels errors aquest model agafa importància degut al fet esmentat pel model anterior en relació a la consistència que ofereix. En el cas del port de Barcelona i de Tarragona no aporta cap millora substancial, però en el cas de València i de l'anàlisi conjunt dels ports BTV, els errors a curt termini augmenten però el model en general es més robust, amb una variació petita a l'augmentar el termini, i que sorprenentment en el cas BTV ha estat fins i tot menor, assolint el valor mínim en aquest model. En les verificacions s'havia vist que efectivament aquest model funcionava molt bé per al període a llarg termini i veiem que estadísticament i en l'anàlisi dels errors és una tendència que es pot confirmar.

Passant al model autoregressiu de segon ordre diferenciat, queda remarcada també la tendència als casos de València i de l'anàlisi conjunt BTV d'assolir major consistència en el model, on en el cas de València s'obté un valor mínim que inclús es inferior al del curt termini. Pels casos de Barcelona i Tarragona l'aportació del model no és suficient com per considerar-lo un model acceptable.

En aquests últims dos models el comportament amb les dades de València i l'anàlisi conjunt BTV ha estat molt similar, amb uns valors d'errors similars i uns estadístics sense pràcticament diferència entre models. Arribat a aquest punt s'ha de fer referència altra vegada al ja esmentat de forma teòrica en la descripció dels models i les seves particularitats. El model que es busca obtenir és el model més simple, de manera que si la diferència és poca entre aquests, no queda justificat l'augment de la complexitat del model. De manera que a l'hora d'escollir un model vàlid entre els dos, aquest seria el model autoregressiu de primer ordre diferenciat.

Finalment, l'últim model ha estat el model de mitjana mòbil corresponent al d'ajustament exponencial simple. Aquest s'allunya de la tendència autoregressiva de les dades, i veiem que no ha aportat millora significativa en cap dels ports, a excepció del cas de l'anàlisi conjunt on si que s'obtenen uns valors acceptables, però en cap cas millor que en els models autoregressius anteriors on ha funcionat millor.

Durant l'anàlisi dels resultats de l'apartat anterior com en l'anàlisi dels errors ha estat interessant observar com han funcionat els models en relació als ports que utilitzàvem. A mode de resum de la discussió es poden extreure els models de major ajust en cada cas.

D'una banda hem vist que per característiques ja explicades de les dades en Tarragona, cap model ha estat vàlid i es pot afirmar que s'hauria de buscar altres solucions per a fer bones prediccions, ja que el model de regressió que hauria funcionat millor té uns valors massa alts pel que fa als errors i no podria ser acceptat.

D'altra banda tenim els altres ports on sí ha estat possible assolir algun model en que les prediccions han funcionat i estadísticament hagin estat acceptables.

El model autoregressiu de primer ordre ha funcionat clarament en el port de Barcelona en ambdós terminis i se'l considera el que millor s'ha ajustat.

En els ports de València i de l'anàlisi conjunt BTV ens trobem que en molts models han compartit el comportament en relació a les conclusions que s'hi han pogut extreure. Això és degut a que són els models amb menors canvis de tendència i que menys punts d'inflexió tenen en les dades, de manera que ha estat motiu suficient per a que aquest comportament envers el model hagi estat similar.

El millor model a curt termini al port de València ha estat el del camí aleatori, però no ha estat robust en l'ampliació del termini de la predicció. Aquest té alguna traça autoregressiva en les correlacions, de manera que per trobar un model ajustat a curt termini i correcte estadísticament es passaria al model autoregressiu de primer ordre. L'altre model vàlid en quant a consistència ja argumentat ha estat el model autoregressiu de primer ordre diferenciat.

Igualment, pel cas de l'anàlisi conjunt dels ports BTV els models que millor s'han ajustat han estat també el model autoregressiu de primer ordre si l'objectiu es centra en prediccions a curt termini, i el model autoregressiu de primer ordre diferenciat si es busca un model més robust als canvis de termini en les prediccions.

6. CONCLUSIONS

Una vegada discutits i analitzats tant els diferents resultats que s'han anat obtenint, podem extreure les conclusions següents:

- La primera conclusió que obtenim és que en general exceptuant el cas de Tarragona, els models més complexos que s'ha analitzat han respost de forma correcta en comparació al model de regressió més simple amb el que estàvem analitzant i comparant.
- Es pot afirmar doncs que els models ARIMA han estat un bon recurs a l'hora de la predicció, i que la varietat que ofereixen ha permès anar ajustant els errors fins a uns límits que han estat prou acceptables, tenint clar que com en tots els models de predicció de sèries temporals, com més en el futur anem més gran és la incertesa i més ha anat creixent aquest error.
- En la línia de la conclusió anterior, es pot afirmar que la separació en dos períodes de verificació ha estat de gran ajuda per observar aquesta característica, i també per veure en quins marges d'error ens estem movent quan augmentem el termini de les nostres prediccions.
- Una altra conclusió que s'ha vist és que no hi ha un model únic que ens hagi funcionat de forma completa en tots els casos, cada conjunt de dades és diferent i ha de ser analitzat de forma independent.
- Lligat amb la conclusió anterior, podem dir que en els ports analitzats la tendència que ha predominat més és la autoregressiva, i això ha donat lloc a que els models que per norma general han donat millor resposta han estat els que tenien termes autoregressius.
- S'ha observat també que en molts casos el model òptim s'ha de determinar d'acord als objectius i a les finalitats que en vulguem extreure de les prediccions. S'ha obtingut models molt acurats a curt termini però amb poca consistència quan l'augmentem, i d'altres que tot i tenir errors una mica superiors, són molt robustos respecte els canvis de període i es mantenen fiables en el temps.
- S'ha vist que el punt feble clar dels models ARIMA que s'ha estudiat i analitzat són clarament els "*turning points*" o punts amb un fort canvi de tendència com ara punts d'inflexió. Si bé és cert que s'havia comentat i explicat en la part teòrica, el cas de Tarragona ens ha permès estudiar de forma pràctica els seus efectes.
- Seguint amb el raonament anterior, l'explicació de que sigui el cas de Tarragona el que menys ha funcionat és degut a que els altres casos han seguit un patronatge similar al procés de containerització, mentre que en Tarragona els canvis en el nombre de TEU han estat molt més bruscs tant en l'augment d'aquests, com en la posterior davallada posterior a la crisi econòmica, que si bé es cert que en tots els casos es pot observar una petita davallada, en el cas de Tarragona el descens també és notori.

- El fet d'estudiar en profunditat els models ARIMA permet lligar la pràctica amb la teoria i elaborar un llistat de pautes o consideracions que permetran elaborar un model amb bon ajust a nivell estadístic, amb paràmetres i gràfics com els que es poden observar en l'[Il·lustració 5](#). Es pot veure del fet que tots els models que han funcionat d'acord als objectius i hipòtesis plantejats prèviament a la seva utilització eren estadísticament correctes, però també és cert que s'han vist models que no s'han rebutjat estadísticament però que després han fallat en les prediccions. Amb tot, això va referir a que s'ha de tenir cautela de no pensar que aquestes pautes donaran el millor model, però sí són una bona eina pràctica quan tractem amb models ARIMA:

- Si l'autocorrelació és positiva al primer *lag*, en general significa que el model s'ajustarà més si s'afegeix un terme autoregressiu.
- Si l'autocorrelació és negativa al primer *lag*, en general significa que el model s'ajustarà més si s'afegeix un terme de mitjana mòbil.
- Un model amb tendència autoregressiva és fàcilment identificable als gràfics d'autocorrelacions, si pel que fa al ACF presenta una caiguda lenta i el PACF presenta un canvi brusc en els primers *lags*, amb autocorrelacions totes positives.
- Un model amb tendència a mitjanes mòbils presentarà un ACF amb canvis bruscs als primers *lags*, i una caiguda lenta al PACF, amb autocorrelacions negatives.
- Si el gràfic de residuals no mostra un patronatge aparentment de valors aleatoris, la sèrie necessita d'una transformació en forma de diferenciació.
- Si el gràfic de residuals mostra un patronatge aleatori però els residuals s'allunyen de la mitjana en el zero, pot ser que el model necessiti d'una diferenciació o que alguna de les consideracions a l'hora de l'elecció del model no hagi estat encertada, pot necessitar d'una revisió en les hipòtesis o raonaments que han portat a la seva elecció.
- En general, les sèries sense diferenciació tendeixen a tenir autocorrelacions positives, i, per tant, certa tendència autoregressiva.
- En general, a l'augmentar el nombre de diferenciacions les autocorrelacions es van fent negatives i el model agafa cada cop més tendència de mitjana mòbil.
- Si s'han aplicat transformacions a les dades, és necessari desfer-les amb anterioritat al càlcul dels intervals de confiança.
- S'ha de tenir en compte que per norma general com més en el futur predim, més gran és la incertesa i més grans els intervals de confiança, que creixen seguint l'anomenada "regla de l'arrel quadrada del temps" ja comentada en la part teòrica.

- Com a conclusió global sobre si els mètodes quantitius d'anàlisi de sèries temporals que hem utilitzat són adequats, seria necessari comparar-los amb els altres mètodes de predicció no tractats en aquesta memòria, com són els qualitius, causals i altres mètodes que utilitzin més d'una variable, com poden ser els que inclouen indicadors econòmics o poblacionals. Dins d'aquesta comparació també seria interessant afegir el punt de vista composicional de les dades, per tal d'obtenir relacions entre variables dins de diferents subconjunts

de les dades que ens facin més explicatius els models (Pawlowsky-Glahn et al, 2015). Això ens podria donar una visió més de conjunt i ens permetria veure si els mètodes univariants tractats són una alternativa fiable i amb una precisió prou fina com per ser complementaris a tots els anteriors.

- Com a punt de partida per a possibles estudis futurs, es pot concloure que seria interessant un estudi amb models que tinguin en compte els factors estacionals com ara els models SARIMA (*seasonal auto-regressive integrated moving average models*), i veure com varien aquestes prediccions i si són més consistents davant de canvis forts de tendència. Aquests estudis haurien de ser a curt termini degut als pocs anys dels que es disposa informació mes a mes, i es podrien comparar amb els mateixos models utilitzats en aquest treball però utilitzant dades solament d'un breu període anterior de temps, permetent-nos fer la comparativa de les dos situacions.

- Finalment, per últim però no menys important, cal destacar que en el global dels ports utilitzats la tendència futura que indiquen les prediccions són de creixement en el nombre de TEU, el qual és important de cara al creixement i desenvolupament de la zona del mediterrani, fent patent i visible la seva potencialitat.

7. BIBLIOGRAFIA

Allix, Y. (2006). De l'atomisation à l'oligopolisation: strategies de concentration des industries maritimes et portuaires conteneurisées. Ecole de Management de Normandie. WP-43/2006.

Autoritat Portuària de Barcelona. Portdebarcelona.cat. Recuperat de http://www.portdebarcelona.cat/home_apb [Última visita 14/8/2017]

Autoritat Portuària de Tarragona. Porttarragona.cat. Recuperat de <https://www.porttarragona.cat/ca/> [Última visita 14/8/2017]

Autoritat Portuària de València. Valenciaport.com. Recuperat de <https://www.valenciaport.com/va/> [Última visita 14/8/2017]

Devore, J. (2016). *Probabilidad y estadística para ingeniería y ciencias*. México D.F.: Cengage Learning.

Glosario de Términos Marítimos Portuarios. (2010). El Salvador. Recuperat de <http://www.amp.gob.sv/ampweb/wp-content/uploads/2016/10/glosario-terminos-maritimos-portuarios-12-08-2010.pdf> [Última visita 8/9/2017]

Guerrero, D., & Rodrigue, J. (2014). The waves of containerization: shifts in global maritime transportation. *Journal Of Transport Geography*, 34, 151-164. <http://dx.doi.org/10.1016/j.jtrangeo.2013.12.003>

Haralambides, H. (2002). Competition, Excess Capacity, and the Pricing of Port Infrastructure. *International Journal Of Maritime Economics*, 4(4), 323-347. <http://dx.doi.org/10.1057/palgrave.ijme.9100053>

Indra, V., Noteboom, T., Parola, F., Satta, G., & Persico, L. (2015). *Port Traffic Forecasting Tool*. Recuperat de <http://www.portopia.eu/wp-content/uploads/2016/07/D1.3-Port-Traffic-Forecasting-Tool.pdf> [Última visita 20/7/2017]

MathWorks - Makers of MATLAB and Simulink. Es.mathworks.com. Recuperat de <https://es.mathworks.com/> [Última visita 3/8/2017]

McCalla, R. (1999). Global change, local pain: intermodal seaport terminals and their service areas. *Journal Of Transport Geography*, 7(4), 247-254. [http://dx.doi.org/10.1016/s0966-6923\(99\)00017-4](http://dx.doi.org/10.1016/s0966-6923(99)00017-4)

Nau, R. (2014). Forecasting with moving averages. Recuperat de https://people.duke.edu/~rnau/Notes_on_forecasting_with_moving_averages--Robert_Nau.pdf [Última visita 14/8/2017]

Nau, R. (2014). Notes on linear regression analysis. Recuperat de https://people.duke.edu/~rnau/Notes_on_linear_regression_analysis--Robert_Nau.pdf [Última visita 14/8/2017]

Nau, R. (2014). Notes on nonseasonal ARIMA models. Recuperat de https://people.duke.edu/~rnau/Notes_on_nonseasonal_ARIMA_models--Robert_Nau.pdf [Última visita 14/8/2017]

Nau, R. (2014). Notes on the random walk model. Recuperat de https://people.duke.edu/~rnau/Notes_on_the_random_walk_model--Robert_Nau.pdf [Última visita 14/8/2017]

Nau, R. (2014). Principles and risks of forecasting. Recuperat de https://people.duke.edu/~rnau/Principles_and_risks_of_forecasting--Robert_Nau.pdf [Última visita 14/8/2017]

Nau, R. (2014). Review of basic statistics and the simplest forecasting model: the sample mean. Recuperat de https://people.duke.edu/~rnau/Review_of_basic_statistics_and_the_mean_model_for_forecasting--Robert_Nau.pdf [Última visita 14/8/2017]

Notteboom, T. (2010). Concentration and the formation of multi-port gateway regions in the European container port system: an update. *Journal Of Transport Geography*, 18(4), 567-583. <http://dx.doi.org/10.1016/j.jtrangeo.2010.03.003>

Notteboom, T., & Rodrigue, J. (2005). Port regionalization: towards a new phase in port development. *Maritime Policy & Management*, 32(3), 297-313. <http://dx.doi.org/10.1080/03088830500139885>

Orellana, L. (2008). Regresión Lineal Simple. Recuperat de http://www.dm.uba.ar/materias/estadistica_Q/2011/1/clase%20regresion%20simple.pdf [Última visita 10/8/2017]

Pallis, A. / Vaggelas, G.K. (2005). Methods for measuring public and private benefits from port services provision: a comparative study. *IAME 2005 Conference*. Limassol.

Pawlowsky-Glahn, V., Egozcue, J., & Tolosana-Delgado, R. (2015). *Modeling and Analysis of Compositional Data*. New York, NY: John Wiley & Sons

Peng, W., & Chu, C. (2009). A comparison of univariate methods for forecasting container throughput volumes. *Mathematical And Computer Modelling*, 50(7-8), 1045-1057. <http://dx.doi.org/10.1016/j.mcm.2009.05.027>

Puertos del Estado. Puertos.es. Recuperat de <http://www.puertos.es/es-es> [Última visita 14/8/2017]

Real Decret Legislatiu 2/2011. (2011). boe.es. Recuperat de <https://www.boe.es/buscar/doc.php?id=BOE-A-2011-16467> [Última visita 14/8/2017]

Real Decret-Llei 8/2017. (2017). Boe.es. Recuperat de https://www.boe.es/diario_boe/txt.php?id=BOE-A-2017-5270 [Última visita 14/8/2017]

Rimmer, P. (2007). Port dynamics since 1965: Past patterns, current conditions and future directions. *Journal Of International Logistics And Trade*, 5(1), 75-97. <http://dx.doi.org/10.24006/jilt.2007.5.1.005>

Robinson, R. (2002). Ports as elements in value-driven chain systems: the new paradigm. *Maritime Policy & Management*, 29(3), 241-255. <http://dx.doi.org/10.1080/03088830210132623>

Sauri Marcha, S., & Robuste Anton, F. (2007). *Modelizaci3n y regulaci3n 3ptima de las concesiones de terminales portuarias de contenedores*. Barcelona: Ca2tedra Abertis.

Song, D. (2003). Port co-opetition in concept and practice. *Maritime Policy & Management*, 30(1), 29-44. <http://dx.doi.org/10.1080/0308883032000051612>

Talley, W. (2006). An Economic Theory of the Port. *Research In Transportation Economics*, 16, 43-65. [http://dx.doi.org/10.1016/s0739-8859\(06\)16003-5](http://dx.doi.org/10.1016/s0739-8859(06)16003-5)

Talley, W. (2007). Chapter 22 Port Performance: An Economics Perspective. *Research In Transportation Economics*, 17, 499-516. [http://dx.doi.org/10.1016/s0739-8859\(06\)17022-5](http://dx.doi.org/10.1016/s0739-8859(06)17022-5)

Tongzon, J., & Heng, W. (2005). Port privatization, efficiency and competitiveness: Some empirical evidence from container ports (terminals). *Transportation Research Part A: Policy And Practice*, 39(5), 405-424. <http://dx.doi.org/10.1016/j.tra.2005.02.001>

Van Klink, H., & Van den Berg, G. (1998). Gateways and intermodalism. *Journal Of Transport Geography*, 6(1), 1-9. [http://dx.doi.org/10.1016/s0966-6923\(97\)00035-5](http://dx.doi.org/10.1016/s0966-6923(97)00035-5)

Van Thai, V., & Grewal, D. (2005). An analysis of the efficiency and competitiveness of Vietnamese port system. *Asia Pacific Journal Of Marketing And Logistics*, 17(1), 3-31. <http://dx.doi.org/10.1108/13555850510672269>